UNDERSTANDING COMPLEX SYSTEMS

Springer COMPLEXITY

Viktor K. Jirsa
A. R. McIntosh
Editors

Handbook on Brain Connectivity

Springer

Springer Complexity

Springer Complexity is an interdisciplinary program publishing the best research and academic-level teaching on both fundamental and applied aspects of complex systems – cutting across all traditional disciplines of the natural and life sciences, engineering, economics, medicine, neuroscience, social and computer science.

Complex Systems are systems that comprise many interacting parts with the ability to generate a new quality of macroscopic collective behavior the manifestations of which are the spontaneous formation of distinctive temporal, spatial or functional structures. Models of such systems can be successfully mapped onto quite diverse "real-life" situations like the climate, the coherent emission of light from lasers, chemical reaction-diffusion systems, biological cellular networks, the dynamics of stock markets and of the internet, earthquake statistics and prediction, freeway traffic, the human brain, or the formation of opinions in social systems, to name just some of the popular applications.

Although their scope and methodologies overlap somewhat, one can distinguish the following main concepts and tools: self-organization, nonlinear dynamics, synergetics, turbulence, dynamical systems, catastrophes, instabilities, stochastic processes, chaos, graphs and networks, cellular automata, adaptive systems, genetic algorithms and computational intelligence.

The two major book publication platforms of the Springer Complexity program are the monograph series "Understanding Complex Systems" focusing on the various applications of complexity, and the "Springer Series in Synergetics", which is devoted to the quantitative theoretical and methodological foundations. In addition to the books in these two core series, the program also incorporates individual titles ranging from textbooks to major reference works.

Understanding Complex Systems

Founding Editor: J.A. Scott Kelso

Future scientific and technological developments in many fields will necessarily depend upon coming to grips with complex systems. Such systems are complex in both their composition – typically many different kinds of components interacting simultaneously and nonlinearly with each other and their environments on multiple levels – and in the rich diversity of behavior of which they are capable.

The Springer Series in Understanding Complex Systems series (UCS) promotes new strategies and paradigms for understanding and realizing applications of complex systems research in a wide variety of fields and endeavors. UCS is explicitly transdisciplinary. It has three main goals: First, to elaborate the concepts, methods and tools of complex systems at all levels of description and in all scientific fields, especially newly emerging areas within the life, social, behavioral, economic, neuro- and cognitive sciences (and derivatives thereof); second, to encourage novel applications of these ideas in various fields of engineering and computation such as robotics, nano-technology and informatics; third, to provide a single forum within which commonalities and differences in the workings of complex systems may be discerned, hence leading to deeper insight and understanding.

UCS will publish monographs, lecture notes and selected edited contributions aimed at communicating new findings to a large multidisciplinary audience.

Viktor K Jirsa · AR McIntosh (Eds.)

With 158 Figures and 3 Tables

Dr. Viktor K Jirsa

Theoretical Neuroscience Group
UMR 6152 Institut Mouvement
& Perception, CNRS
163 av. de Luminy, CP910
F-13288 Marseille cedex 09
France
viktor.jirsa@univmed.fr
and
Center for Complex Systems & Brain
Sciences, Physics Department,
Florida Atlantic University
Boca Raton, FL33431
USA

Dr. AR McIntosh

Rotman Research Institute
Baycrest Centre
University of Toronto
CIHR Scientist
3560 Bathurst Street
Toronto, Ontario, M6A 2E1
Canada
rmcintosh@rotman-baycrest.on.ca

Library of Congress Control Number: 2007926434

ISSN 1860-0832
ISBN 978-3-540-71462-0 Springer Berlin Heidelberg New York

Typesetting: by the author and Integra, India using a Springer LaTeX macro package
Cover design: WMX Design, Heidelberg

Printed on acid-free paper SPIN: 11539926 5 4 3 2 1 0

Preface

Neuroanatomical work of Golgi and Ramon y Cajal set the stage for the notion that nervous system operations arose from the interplay amongst its parts. While Golgi considered the brain as a continuous system (reticular), Cajal proposed what became known as the neuron doctrine, in which the elemantary unit of brain function, the neuron, was an separable entity, but whose operations were dependent on its input from other neurons. These critical ideas grew in the company of emerging theories of brain function that emphasized the need for interaction between parts. In the early 1940s Warren McCulloch and Walter Pitts gave birth to modern neural network theories by proposing brain function to arise from the network action of binary neurons. Their network dynamics, though simple, resulted in remarkably general computational feats. Reinforcing the concept of network action, in his 1949 book *Organization of behavior* Donald Hebb proposed the modification of neural connections to be a function of their systematic usage and through such laid the basis for one of the most commonly used learning rules in neural nets. However, it is probably fair to say that the McCulloch-Pitts network model finds its place more appropriately in the lineage of artificial neural networks than in the understanding of the dynamics of biological systems. In subsequent developments, McCulloch-Pitts systems formed the basis for the two-layer "perceptrons" of Rosenblatt (1958), the symmetric Hopfield (1982) networks and Grossberg's Adaptive Resonance Theory (ART) (1976), which extended the computational properties to permit object categorization, competitie learning and content-addressable memory (i.e. the system correctly yields an entire memory from any subpart of sufficient size). Parallel developments in the 1940s find their roots in the fields of electronic engineering and control system theory and culminate in the field of Cybernetics pioneered by Norbert Wiener. Cybernetics emphasizes regulatory feedback in living and non-living systems and as such connected functional and neural architectures in the field of neuroscience. Here architectures are composed of functionally defined interconnected components giving rise to teleological mechanisms and emphasizing the purpose of a process or behavior. The notions developed in

Cybernetics are closely interwoven with the thinking of other disciplines related to brain function, including radical constructivism in philosophy and autopoiesis (literally "self-creation") in neurobiology advanced by Heinz von Foerster, Humberto Maturana and Francisco Varela among others. In the late 1960s, Cybernetics as a term was increasingly replaced by Self-organization and in the 1970s and 80s by Complex System Theory. One of the reasons is certainly the adoption of these notions by physicists and its consequent developments in pattern formation, nonlinear dynamics and non-equilibrium systems. Hermann Haken pioneered many of these concepts and summarized these in his seminal paper on Synergetics in Reviews of Modern Physics in 1975.

Empirical evaluation of brain networks took somewhat longer to develop. Though the idea of large-scale network operations was at the base of early ideas of nervous system operations (e.g., Bethe's chain theory, Lashley's ideas of mass action, Hebb's cell assemblies), the technology was insufficient to really capture network dynamics in any quantifiable means. As physiological data were moved from paper storage to digital format, analytic methods sensitive to the dependencies of firing arose. Seminal works of Gerstein in 1970 and later with Aertsen in the late 1980's provide the mathematical and theoretical foundation for the ideas of *functional* and *effective connectivity* – ideas that now are "household" terms in neuroscience. In the same period, the pioneering work of Gevins and colleagues, published in Science, showed the interplay of cortical sources measured with EEG in humans could be directly related to different stages in cognitive operations. While these works inspired continued development of network analytic methods in electrophysiology, the measure of brain network interactions has been most prolific in functional neuroimaging. Starting first with SPECT and PET, and then functional MRI, there was a new opportunity for comprehensive measures from most of the brain in persons performing a wide range of mental functions. Such data quantified distributed action to be brought to bear on network theories of brain operation. Authors in the current volume were among the first to explore network operations revealed in functional neuroimaging (Friston, Horwitz, McIntosh). The current state of the field has entire research programs geared specifically to methods of measuring brain network dynamics, with parallel advances in the quantification of the anatomical connectivity.

Contemporary brain theories are deeply rooted in the ideas of dynamic systems, self-organization and large scale connectivity. It is posited that brain function arises from the interaction of connected brain areas, in which functional differentiation is performed by the more peripheral areas, whereas the ensuing cognitive integration rests heavily on the large scale network action. The presence of local and global elements involved in brain dynamics poses novel constraints, unknown to the dynamics and pattern formation mechanisms encountered in physical and chemical systems. To uncover the mechanisms of brain function in networks with various local and global architectures is one of the major challenges we are facing these days.

Our Handbook arose from the fourth in a series of annual meetings focused on brain connectivity (http://www.hirnforschung.net/bcw/). These intense workshops brought together anatomists, physiologists, cognitive and, computational neuroscientists to discuss the current issues on how to integrate brain connectivity across structural and functional domains. Motivated by the discussions at this and the prior meetings, here we wish to provide an account of the current knowledge on the imaging, analysis and theory of the anatomical and functional connectivity of the brain.

The Handbook on Brain Connectivity would not have been possible without the outstanding contributions of the Handbook authors. All contributors are leading experts in various fields concerning structural and functional brain connectivity. In the first part of the Handbook, the chapters focus upon an introduction and discussion of the principles underlying connected systems in the brain. The second part introduces the currently available non-invasive technologies for measuring structural and functional connectivity in the brain. Part three provides an overview of the analysis techniques currently available and highlights new developments. Part four introduces the application and transfer of the concepts of brain connectivity to behavior, cognition and the clinical domain. We also wish to thank Arpan Banerjee, Ajay Pillai, Murad Qubbaj, Young-Ah Rho, Roxana Stefanescu and Maria Tassopoulos for their tireless editorial help. We are most grateful to Technical Editor Thomas Ditzinger at Springer Pulisher, who has been always available for guidance and advice during the creation of the Handbook.

Viktor K Jirsa
Anthony R McIntosh

Contents

Part I

Principles Underlying Structural and Functional Brain Connectivity

Neuronal Dynamics and Brain Connectivity

Michael Breakspear[1,2] and Viktor K Jirsa[3,4]

[1] School of Psychiatry, University of New South Wales, and The Black Dog Institute, Randwick, NSW, 2031, Australia.
[2] School of Physics, University of Sydney, NSW 2006, Australia
[3] Theoretical Neuroscience Group (CNRS), UMR6152 Mouvement & Perception, 13288 Marseille, France
[4] Center for Complex Systems & Brain Sciences, Physics Department, Florida Atlantic University, Boca Raton FL33431, USA

The fluid nature of perceptual experience and the transient repetition of patterns in neurophysiological data attest to the dynamical character of neural activity. An approach to neuroscience that starts from this premise holds the potential to unite neuronal connectivity and brain activity by treating space and time in the same framework. That is the philosophy of this chapter. Our goals are threefold: Firstly, we discuss the formalism that is at the heart of all dynamical sciences, namely the *evolution equation*. Such an expression ties the temporal unfolding of a system to its physical properties and is typically a differential equation. The form of this equation depends on whether time and space are treated as continuous or discrete entities. Secondly, we aim to motivate, illustrate and provide definitions for the language of dynamical systems theory - that is, the theoretical framework that integrates analysis and geometry, hence permitting the qualitative understanding and quantitative analysis of evolution equations. To this end we provide a mini-encyclopedia of the basic terms of phase space analysis and a description of the basic bifurcations of dynamics systems. Our third aim is to provide a survey of single neuron and network models from a historical and pedagogical perspective. Here we first trace microscopic models from their birth in the 1950's showing how the neuronal firing properties can be understood as a bifurcation in the underlying phase space. Then we review the spatiotemporal network dynamics, which emerges as a function of the networks anatomical connectivity.

Introduction: Dynamics and the Brain

The firing of a neuron subsequent to an increase in synaptic input is a crucial neuronal event that is best understood from a dynamic system perspective. Whilst statistical techniques are crucial to the detection of synchrony and

change in neuroscience data, the study of dynamics uniquely permits an understanding of their causes. "Evolution" equations – which embody a system's dynamics - form the basis of all major theories in the physical sciences, from Newton's $F = ma$ to Schrödinger's wave equation and Maxwell's electromagnetic theory. There is no reason to believe that mathematical formalisms of neuronal dynamics won't eventually underpin and unify neuroscience. Indeed, over recent decades, dynamical formulations of brain activity have become sufficiently advanced to give rough outline to a "unified theory of brain dynamics". Such a theory will also inform studies of brain connectivity.

What is the origin of the brain's dynamic character? During the 20^{th} century, extraordinary progress was made in elucidating basic neurophysiological processes and their role in neural phenomena such as neuronal firing and action potential propagation. Incorporating these processes into a set of evolution equations yielded quantitatively accurate spikes and thresholds, leading to the Nobel prize for Hodgkin and Huxley. These equations are based upon the physical properties of cell membranes and the ion currents passing through transmembrane proteins. Extending this theory from a patch of cell membrane to whole neurons and thence to populations of neurons in order to predict macroscopic signals such as the electroencephalogram (EEG) is a dominant focus in this field today. Linking neuronal dynamics to theories of cognition also remains a major goal.

Dynamics has a spatial as well as a temporal character and this makes it relevant to the subject of this handbook, brain connectivity. It can be argued that all forms of information processing in neuronal systems can be understood as particular types of spatiotemporal dynamics and their bifurcations. With this in mind, our primary objective is to provide a "ground-up" overview of the dynamical approach to neuroscience. We also aim to overview some of the recent developments in this field, such as those that establish a link between statistics and dynamics and proposals that provide putative network-based cognitive mechanisms with a biophysical underpinning. Attempts to employ dynamics to unify neurophysiological phenomena are also covered. Section 4, dealing with macroscopic spatiotemporal dynamics, implicitly incorporates connectivity by way of its joint treatment of space and time.

Section 1 provides an overview of the central concept of dynamics - the "evolution equation" - and reviews the variety of forms that it can assume. In Sect. 2, we overview the mathematical concepts required to understand the behavior of such equations, with an emphasis on a geometric approach. In doing so, we also show how many of the stochastic approaches more familiar to neuroscientists are specific forms of dynamical systems when they satisfy certain stability conditions. In Sect. 3, we provide a taxonomy of key neuronal models – that is, particular forms of neuronal evolution equations, with an emphasis on small scale systems. Section 4 then focuses on large scale neuronal dynamics. We argue that there is a one-to-one relationship between modes of information processing in neuronal systems and their spatiotemporal

dynamics. Likewise, changes between such forms correspond directly with changes in the dynamics, mediated by a bifurcation or similar mechanism. The chapter concludes in Sect. 5 with some of the exciting recent developments in the field of neuronal dynamics and their putative links to other "hot topics" in neuroscience.

1 Evolution Equations: How to Make a Dynamical System

Evolution equations lie at the heart of dynamics. They state how a set of dynamical variables change in accordance with the underlying properties of the system they characterize. The most famous example of an evolution equation is Newton's "second law of mechanics" which describes the acceleration of an object as $\mathbf{F} = m\mathbf{a}$. More technically this is written as,

$$\frac{d\mathbf{v}\,(t)}{dt} = \frac{\mathbf{F}}{m}, \qquad \frac{d\mathbf{x}(t)}{dt} = \mathbf{v}\,(t) \tag{1}$$

where $\mathbf{v(t)}$ is the velocity of an object at position $\mathbf{x}(t)$. The left hand sides (LHSs) of these equations express the temporal derivative – the rate of change of a variable. The right hand sides (RHSs) link these changes to the properties of the system. The goal of calculus is to understand the resulting evolution of these variables as a function of time. In (1), it is possible to find an explicit solution for the evolution of $\mathbf{x}$ in terms of time,

$$\mathbf{x}(t) = \mathbf{x}(0) + \mathbf{v}\,(0)\,t + \frac{\mathbf{F}t^2}{2m}\,, \tag{2}$$

where $\mathbf{x}(0)$ and $\mathbf{v}(0)$ are the 'initial conditions' of $\mathbf{x}$ and $\mathbf{v}$. Equation (2) allows us to know the exact future position of an object given its current state and any applied constant force. We can see that as time increases the RHS of (2) will be dominated by the quadratic term, t^2 so that an object subject to a constant force will be increasingly rapidly displaced. In more complex systems, as encountered in neuroscience, such explicit closed form solutions generally cannot be found. Moreover, their approximations are typically so cumbersome that understanding the nature of the dynamics from such algebraic equations is not straightforward. However, one may gain a deep understanding of the nature of a system's dynamics without relying only on algebraic solutions. This can be achieved through the geometric approach to dynamical systems, outlined in Sect. 2, which unifies algebraic analysis and topology.

The essential requirements for an evolution equation are a set of evolving variables which we denote $\mathbf{Z}(\mathbf{x}, t)$ and a set of system parameters denoted $\mathbf{a}$. The former represent the current states of properties such as transmembrane potentials, neuronal firing rates, extracellular field potentials, as they vary in time t and position $\mathbf{x}$. The parameters $\mathbf{a}$ are those properties which

can be considered as static or change very slowly in comparison to the dynamical variables **Z**. Nernst potentials, conduction velocities and ion channel time constants are typical neural parameters. All of these variables are then combined with a "differential operator" – which introduces the crucial factor of change – and an algebraic expression – which determines how this change relates to the properties of the system – to form an evolution equation.

We now progress through the various forms that such equations can assume, from the simplest to the more complex. Exemplar neuronal models of each system are given in Sect. 3. Further suggested reading is provided where appropriate.

1.1 Difference Maps: Discrete Time and Discrete Space

The simplest form of determining the future state of a dynamical system from its present state is through a *difference map*,

$$\mathbf{Z}\left(t+1\right) = \mathrm{F}_{\mathbf{a}}\left[\mathbf{Z}\left(t\right)\right], \tag{3}$$

where t runs discretely as 0,1,2,... Note that the subscript $\mathbf{a}$ denotes the parameterization of F. The so-called "logistic" equation,

$$\mathrm{F}_{a}\left[Z\left(t\right)\right] = aZ(1-Z), \tag{4}$$

is a very well-known one-dimensional (scalar) example of a difference equation. The evolution of this relatively simple (quadratic) nonlinear equation is illustrated in Fig. 1.

The logistic map, and other simple algebraic forms, has been used extensively to elucidate basic, generic properties of nonlinear dynamics (Collet & Eckmann 1980, Cvitanovic 1984). They can exhibit a rich complexity even when the algebraic equations are simple as in (4). Examples of their use include elucidating the fundamental principles of chaotic dynamics (Guckenheimer 1987) and the transition from regular to chaotic motions (Feigenbaum

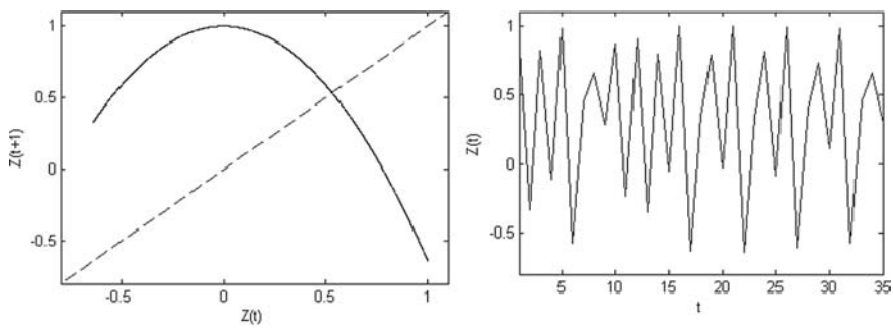


Fig. 1. (**a**) Logistic equation (4) with $a = 1.64$. (**b**) Resulting chaotic time series. Maps of this type have been used to study the basic properties of nonlinear systems. They have a less extensive role in modeling neural systems

1987). These concepts are discussed and illustrated in Sect. 2, below. An excellent introduction to this fascinating field is given by Baker & Gollub (1990).

The spatiotemporal properties of nonlinear dynamics can also be studied within this framework, through the use of *coupled difference maps*,

$$\mathbf{Z}\left(x_i, t+1\right) = \mathrm{F}_{\mathbf{a}}\left(\mathbf{Z}\left(x_i, t\right), \sum_{j \neq i} \mathrm{H}_c\left[\mathbf{Z}\left(x_j, t\right)\right]\right) \tag{5}$$

where x_i denotes the spatial position of the i-th subsystem. The "coupling function" H_c introduces the activity from all other nodes into the dynamics of this node. The subscript c denotes the strength of the coupling influence and is traditionally normalized so that $0 \leq c \leq 1$. Hence, if F embodies local neural dynamics, H incorporates the spatial characteristics of synaptic connectivity. Just as autonomous difference maps can be used to elucidate basic dynamical principles, coupled difference maps permit an understanding of the fundamentals of dynamic synchronization (Maistrenko et al. 1998). Often the influence of the local versus global dynamics can be linearly partitioned as,

$$\mathbf{Z}\left(x_i, t+1\right) = \mathrm{F}_{\mathbf{a}}\left[\mathbf{Z}\left(x_i, t\right)\right] + \sum_{j \neq i} \mathrm{H}_c\left[\mathbf{Z}\left(x_j, t\right)\right]. \tag{6}$$

A fascinating, early example of a coupled difference-map neural model is that of McCulloch & Pitts (1943) which we discuss in Sect. 3. However, because of the discrete nature of time in difference maps, and the fact that their study has been characterized by using very basic algebraic expressions, they rarely figure in biophysical models of neural systems. On the other hand, they have been used extensively to study the basic properties of high dimensional nonlinear dynamics (Kaneko 1997), including the onset of synchronization amongst two or more subsystems (Ashwin et al 1997). Put another way, they are mathematically pleasing because they permit an analytic understanding of the universal principles of dynamics and synchronization, but limited in their value to neuroscientists because their simplicity prohibits one from identifying the relative contribution of particular physiological processes to specific dynamical behaviors.

1.2 Ordinary Differential Equations: Continuous Time and Discrete Space

One obvious step towards physiological realism is to make time continuous! This can be achieved by exploring neural systems, whose evolution is governed by an ordinary differential equation (ODE),

$$\frac{d\mathbf{Z}\left(t\right)}{dt} = \mathrm{F}_{\mathbf{a}}\left[\mathbf{Z}\left(t\right)\right]. \tag{7}$$

where as above $\mathbf{Z}(t)$ is a set of dynamical variables. This is the form of equation for most traditional neuronal models such as the Hodgkin Huxley

model in which case $\boldsymbol{Z}_1 = V$ is the transmembrane potential and F takes the form,

$$\mathrm{F}_{\mathbf{a}}\left[V\left(t\right)\right] = \sum_{ion} f_{ion}\left[V\left(t\right)\right] + I. \tag{8}$$

The summation on the RHS is taken over all ion channels. For each ion species

$$\frac{df_{ion}\left(t\right)}{dt} = g_{ion}\left[V\left(t\right)\right] \tag{9}$$

represents the dynamics of the local voltage-dependent channel currents for each ion species. I represents synaptic currents which flow through ligand-gated channels or via an experimentally introduced electrode. As with difference equations, spatiotemporal dynamics are achieved by employing a coupling function H_c to introduce interdependence between systems,

$$\frac{d\mathbf{Z}\left(x_i,t\right)}{dt} = \mathrm{F}_{\mathbf{a}}\left(\mathbf{Z}\left(x_i,t\right), \sum_{j\neq i} \mathrm{H}_c\left[\mathbf{Z}\left(x_j,t\right)\right]\right). \tag{10}$$

Hence, if (8) models the dynamics of a single neural system, (10) adds the interaction between two or more systems, creating a dynamic neural network. The ensemble is spatially discrete with a finite number N of subsystems so that the subscript indices $i, j = 1, 2, \ldots, N$. As with coupled difference equations, it is often possible to bipartition the influence of the local and distant terms in (10) as

$$\frac{d\mathbf{Z}\left(x_i,t\right)}{dt} = \mathrm{F}_{\mathbf{a}}\left(\mathbf{Z}\left(x_i,t\right)\right) + \sum_{j\neq i} \mathrm{H}_c\left[\mathbf{Z}\left(x_j,t\right)\right]. \tag{11}$$

Such is the case when local recurrent axons and long-range afferents each project onto separate classes of neurons. In this case the long-range afferents are modeled as acting, through ligand-gated ion channels, via the synaptic currents. Hence,

$$\frac{d\mathbf{Z}\left(x_i,t\right)}{dt} = \sum_{ion} f_{ion}\left[\mathbf{Z}\left(x_i,t\right)\right] + I\left(x_i,t\right), \tag{12}$$

where the induced synaptic currents,

$$I\left(x_i,t\right) = \sum_{j} \mathrm{H}_c\left[\mathbf{Z}\left(x_j, t-\tau_j\right)\right] + I_{external}, \tag{13}$$

introduce the afferent inputs from other systems $\mathbf{Z}(x_j,t)$ that arrive after a time delay τ_j - permitting finite speed axonal conduction. Because space is discrete, the time delays are also discretely distributed. Differential equations with time delays are treated thoroughly in the Chapter by Campbell. We only introduce them here because they are important in the conceptual transition from discrete to continuous space to which we now turn. A review of neuronal synchrony as modeled by coupled ODE's is provided by Breakspear (2004).

1.3 Integrodifferential Equations: Continuous Time and Continuous Space

In the case where we wish to model the inputs to region x_i arising from a continuously distributed neuronal ensemble, we integrate the afferent induced currents (13) continuously over space,

$$I(x_i,t) = \int_{\Omega}\int_{-\infty}^{t} h(\mathbf{x}-\mathbf{x}')\,\mathrm{H}\left[\mathbf{Z}(\mathbf{x}-\mathbf{x}',t-t')\right]dt'd\mathbf{x}', \tag{14}$$

where the spatial integration d$\mathbf{x}$' is taken over the spatial domain Ω of the neural system. Note that this also requires that we integrate over the (now) continuously distributed time delays, t'. We have also partitioned the coupling function into two parts, H and h. H determines which variables from any given system enter into the inter-system coupling, and how they do so. Typically H itself has two components, an "activation" function that converts local membrane potentials of the distant systems into firing rates ρ - which then propagate outwards - and synaptic kernels η which model how these propagating action potentials influence post-synaptic potentials as they arrive,

$$\mathrm{H}\left[\mathbf{Z}(\mathbf{x},t)\right] = \eta(t)\,\rho\left(\mathbf{Z}(\mathbf{x},t)\right) \tag{15}$$

Specific forms of η and ρ are provided in Sect. 4. The coupling function h captures the spatial dependency of the strength of the afferent inputs. This function is also known as the 'synaptic footprint' (Coombes 2003) because it reflects the nature and density of synaptic connections as they change with the distance from their origin. Substituting the synaptic inputs (14) into the differential (12) and collapsing all local contributions into

$$N\left[\mathbf{Z}(\mathbf{x},t)\right] = \sum_{ion} f_{ion}\left[\mathbf{Z}(\mathbf{x},t)\right] \tag{16}$$

we obtain

$$\frac{d\mathbf{Z}(\mathbf{x},t)}{dt} = N(\mathbf{Z}(\mathbf{x},t)) + \int_{\Omega}\int_{-\infty}^{t} h(\mathbf{x}-\mathbf{x}')\,\mathrm{H}\left(\mathbf{Z}(\mathbf{x}-\mathbf{x}',t-t')\right)dt'd\mathbf{x}', \tag{17}$$

an integrodifferential equation. It may be considered a general form of a neural mass model because the exact nature of the synaptic "footprint", the activation function and the synaptic kernels remain unspecified. For example within this framework, it would be possible to use the precise form of the lateral inhibition that has been shown to allow sensory networks to be inherently tuned to particular spatial frequencies (Ratliff et al. 1969).

1.4 Partial Differential Equations: Continuous Time and Continuous Space but Constrained Connectivity

In some contexts, it may be preferable to express (17) with spatial and temporal derivatives only - rather than a combination of temporal derivatives with spatial and temporal integrations. Such differential representation is useful if the connectivity function h is sufficiently simple, smooth and translationally invariant, because then only a few spatial derivatives are needed to capture the connectivity. For example, given appropriate forms of h and H (see sect. 4) (17) can be rewritten as a partial differential equation of the form,

$$\frac{\partial^2 \mathbf{Z}(\mathbf{x},t)}{\partial^2 t} + a\frac{\partial \mathbf{Z}(\mathbf{x},t)}{\partial t} + b\frac{\partial^2 \mathbf{Z}(\mathbf{x},t)}{\partial^2 x} + c\mathbf{Z}(\mathbf{x},t) = (d + \frac{\partial}{\partial t})\rho(\mathbf{Z}(\mathbf{x},t)) \quad (18)$$

The coefficients $a,\ b,\ c$ and d depend on system parameters such as conduction velocities and the synaptic footprint parameter σ. Such an equation, expressing the evolution of neuronal systems, continuously in space and time, but with specific types of connectivity was first derived for macroscopic neuronal dynamics by Jirsa and Haken (1996, 1997) and Robinson et al. (1997). Pioneering work that led to this formulation started as early as the 1970s (Wilson 1973, Wilson & Cowan 1973, Nunez 1974, van Rotterdam et al. 1982). Comparing (11) and (18) we see that in the former, spatial coupling is introduced explicitly through the second term on the right hand side. In the latter, space enters the temporal dynamics through the (second order) spatial derivative on the left hand side. However, under certain conditions these two approaches can be equivalent.

1.5 Stochastic Differential Equations

All of the equations above capture the dynamical evolution of each of the values $\mathbf{Z}$ of the system of interest. In the case of a microscopic system, these variables may include transmembrane potential and ion channel conductance of a small patch of cell membrane. Evolution equations, such as (11) and (18) may also describe neural systems in the mesoscopic (<mm) and even macroscopic (∼cm) scales. In such cases, the variables of interest represent mean values averaged over the appropriate scales. Such equations are hence known as mean field approximations. Before proceeding further, it is worth describing evolution equations which capture the dynamics for the entire probability distributions $p(\mathbf{x},t)$ rather than just the mean. Such models allow for stochastic inputs to a system which nonetheless obeys deterministic rules. They take the form,

$$\frac{\partial p(\mathbf{x},t)}{\partial t} = \frac{\partial\left((f+s)\,p(\mathbf{x},t)\right)}{\partial x} + \frac{w^2}{2}\frac{\partial^2 p}{\partial x^2} \quad (19)$$

where s represents the (stochastic) inputs to the system and f is the form of the deterministic dynamics. As described in Harrison et al. (2005), the first

term of the RHS describes the evolution of the probability distribution under the influence of the inputs s and the nature of the physiological system f. The second term describes the tendency of the distribution to disperse under the influence of the stochastic elements at rate w.

Whereas (11) and (18) are mean field equations, (19) is an example of a broader class of "neural field" equations, capturing the evolution of the entire probability distribution. There are a number of intriguing reasons to generalize neural evolution equations from mean field formulations to capture the evolution of the entire distributions. For example, consider two neural populations with the same mean membrane potentials, but where the second population has a larger variance. If the mean potential is below the threshold for firing, this difference in variance will imply that a greater proportion of neurons in the second population will be supra-threshold and hence firing (Fig. 2). These neurons, through local feedback, will in turn have a greater effect on the local mean membrane potential, driving it upwards or downwards – depending on whether the local feedback is excitatory or inhibitory. Put alternatively, modeling the entire distribution rather than just the mean permits the higher order moments of the neural states to interact (Harrison et al. 2005).

Solutions to (19) are possible only in very restricted cases. The development of numerical techniques – required to gain important insights into the dynamics – is a very active area of research. One important method in this vein relies upon a "modal decomposition", whereby the entire distribution is truncated to a few low order modes. The very restricted case, reducing such an equation to the first moment – the mean – returns us to the mean

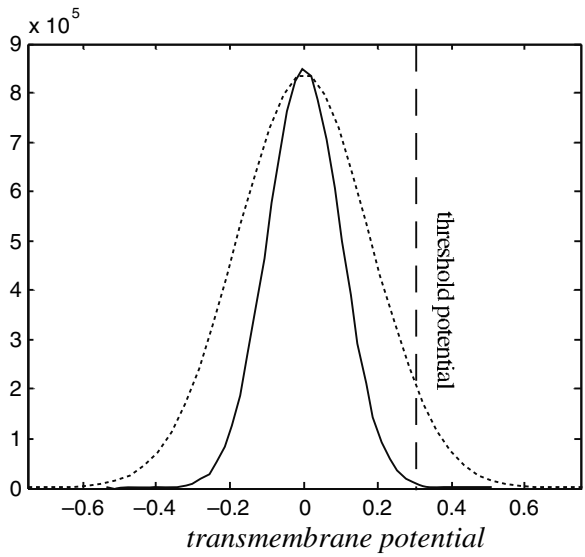


Fig. 2. Two different distributions of neural membrane potentials, with the same mean states, will in general have different mean firing rates (arbitrary units)

field formulations discussed above. In situations where higher order moments of the distribution (variance, kurtosis etc) are retained, the possibility exists for deep, but tractable representations of complex neuronal dynamics. Hence interactions between stochastic and deterministic processes, as embodied in (19) can be formally studied.

2 The Geometry of Dynamics: A Mini-Encyclopedia of Terms

In this section, we step back from a consideration of the forms that evolution equations can take and overview the crucial geometrical means of understanding them.

2.1 Basic Dynamical Concepts

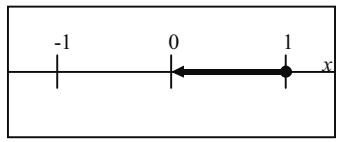
As mentioned in the Introduction, nonlinear differential equations can be notoriously intractable with regards to exact analytic solutions. However, a thorough understanding of their dynamics is very often possible by combining analysis and geometry. In this Section, we provide the central defining terms through the exploration of some simple dynamical systems. In interests of brevity we have sought to explain the intuitive meaning of the terms, keeping technical definitions to a minimum. Most of these terms are given more formal definitions in standard dynamical systems textbooks (e.g. Strogatz 1994). Illustrated examples of all terms follow in subsequent sections.

For any study of geometry, we require a space in which to embed our objects of study. For evolution equations, a ***manifold*** fulfills this purpose. Put simply, a manifold is a space which can be locally stretched or deformed into a Euclidean space whilst having a variety of global shapes. Hence the local structure sustains the intuitive meaning of terms such as a ***neighbourhood*** (a ball of small radius) which are crucial for issues requiring a "distance", well defined for Euclidean space. The global structure of a manifold, on the other hand, can be quite complicated, and may be 'bounded' (like the unit interval) or 'unbounded' (like the Euclidean plane), 'simply connected' (like a sphere) or not simply connected (like a torus). A ***differentiable manifold*** has the additional properties required to support differentiation. The planar surface, a torus and a sphere are differentiable manifolds. Although the properties of these spaces may seem trivial, a formal definition of a differentiable manifold must be able to support quite general dynamical systems. For example, the manifold of a partial differential equation has infinite dimension!

A ***phase space*** is a differentiable manifold whose axes are spanned by the dynamical variables $\mathbf{Z} = \{Z_1, Z_2, Z_3, \ldots\}$ of an evolution equation. The ***topology*** ("shape") of the phase space is chosen to match the properties of these variables. For example, the plane ($\mathrm{R}\times\mathrm{R}$) is a suitable phase space for a system with two membrane potentials. For a system where the two variables

are phases varying between 0 and 2π, the torus $(S \times S)$ is preferable because of the periodic nature of the boundaries.

We can think of a point in phase space as the instantaneous state $\mathbf{Z}(t)$ of our system. If we substitute this state into our evolution equation, we would get the instantaneous rate of change of the system $d\mathbf{Z}(t)/\mathrm{d}t$ when in that state. This defines a ***tangent vector*** in the phase space, telling us how the system will evolve into its next state $\mathbf{Z}(t')$. This critical step – of linking dynamics to geometry – is captured by the ***vector field***, a directed flow through a phase space which embodies the evolution equation. More technically, a vector field assigns a vector to every point in phase space which is precisely the solution of the evolution equation at that point. Hence these vectors capture both the rate and direction of change of the system. For example, the vector field corresponding to the trivial one-dimensional equation $\mathrm{d}x/\mathrm{d}t = -x$ is just the set of all vectors of length x pointing towards the origin. The vector field at $x = 1$ is a vector of length 1 directed towards the origin.



In this way the algebra of the evolution equations and the analysis methods of geometry are linked (Vector fields are often *represented* as arrows overlaid on the phase space but more technically they are defined on a related space called the ***tangent bundle***). An ***orbit*** or ***trajectory*** is a connected path through phase space which is always tangent to the vector field. Hence an orbit traces the time-dependent solution to a dynamical system through a succession of instantaneous states. It captures the manner in which a system will change according to the evolution equation. The starting point of such an orbit is called its ***initial condition***. Examples are given in Fig. 3.

$$\frac{dz_1}{dt} = z_2 \ , \qquad \frac{dz_2}{dt} = \mu\left(1 - z_1^2\right) z_2 - z_1 \ . \tag{20}$$

A time series of this system for $\mu = 2$, showing the periodic nature of the oscillations, is given in Fig. 4(a). A single orbit, commencing with an initial condition in close proximity to the origin is shown in the planar phase space spanned by z_1 and z_2 in Fig. 4(b). It can be seen that this orbit diverges rapidly away from the origin and towards a closed loop in phase space, corresponding to the appearance of periodic oscillations in the time series. The appearance of periodic oscillations in the system motivates us to consider an alternative phase space representation, achieved by a change of coordinates to amplitude A and phase θ,

$$A = \sqrt{z_1^2 + z_2^2} \ , \qquad \phi = \arctan\left({}^{z_2}\!/_{z_1}\right) . \tag{21}$$

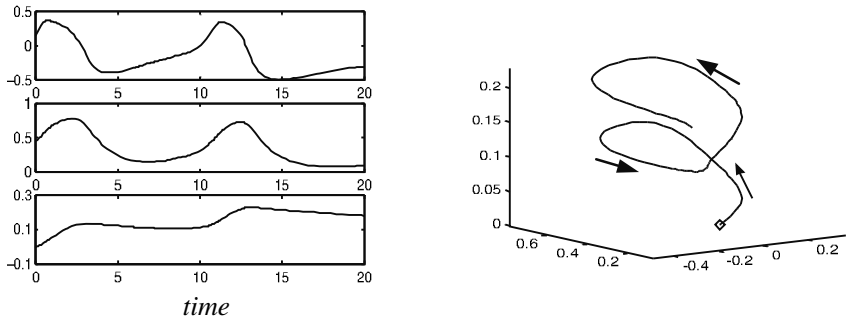


Fig. 3. Example of a time series (left hand panel) of three variables $\mathbf{Z} = \{Z_1, Z_2, Z_3, \dots\}$ of a dynamical system briefly evolving along an orbit in phase space (right hand panel). The diamond represents the initial condition and the arrows represent tangent vectors in phase space

Such a cylindrical manifold, whose periodic boundary conditions embody the nature of the oscillatory dynamics, is shown in Fig. 4c. The convergence of the orbits shown in panels (b) and (c) onto the closed loop, on which they then remain, motivates the concept of an ***invariant set*** of the dynamics. Intuitively, this is simply a set of points (e.g. single point, closed loop, etc.) in which orbits remain once they enter. More formally if $\boldsymbol{F}$ represents the flow of a dynamical system then an invariant set $\boldsymbol{A}$ satisfies $\boldsymbol{F}(\boldsymbol{A}) \subseteq \boldsymbol{A}$. The Van der Pol system in Fig. 4 has two invariant sets, one at the origin and the closed loop as shown. A variety of other orbits, from distinct initial conditions, are shown in panel (d). In each case, the orbits approach the limit cycle.

This simple observation motivates the crucial concept of an ***attractor***, a bounded (i.e. finite) invariant set which is approached by the orbits from a "large set" of initial conditions. Traditionally, a large set implied an "open neighborhood" of the attractor. More recently the concept of an attractor has been generalized to mean any set with a non-zero probability measure (Milnor 1985) meaning that there is a (possibly very small but still non-zero) chance that an orbit from a randomly chosen initial condition will flow onto the attractor. On the other hand, there may be initial conditions arbitrarily close to the attractor that nonetheless flow elsewhere. This distinction is important in the setting of synchronization (Ashwin & Terry 2000) and we explore it further below.

We have hence seen ***fixed point*** and ***limit cycle*** attractors. A ***chaotic attractor*** has already been illustrated for the logistic equation in Fig. 1. In comparison to a limit cycle which endlessly repeats its prior states, a chaotic attractor *never* repeats a state although is nonetheless bounded and invariant. More formally, a chaotic attractor exhibits ***sensitive dependence to initial conditions*** – that is any two orbits, no matter how close initially - diverge at an exponential rate. This rate of divergence is captured by the largest ***characteristic exponent***, which is positive for a chaotic attractor. In contrast,

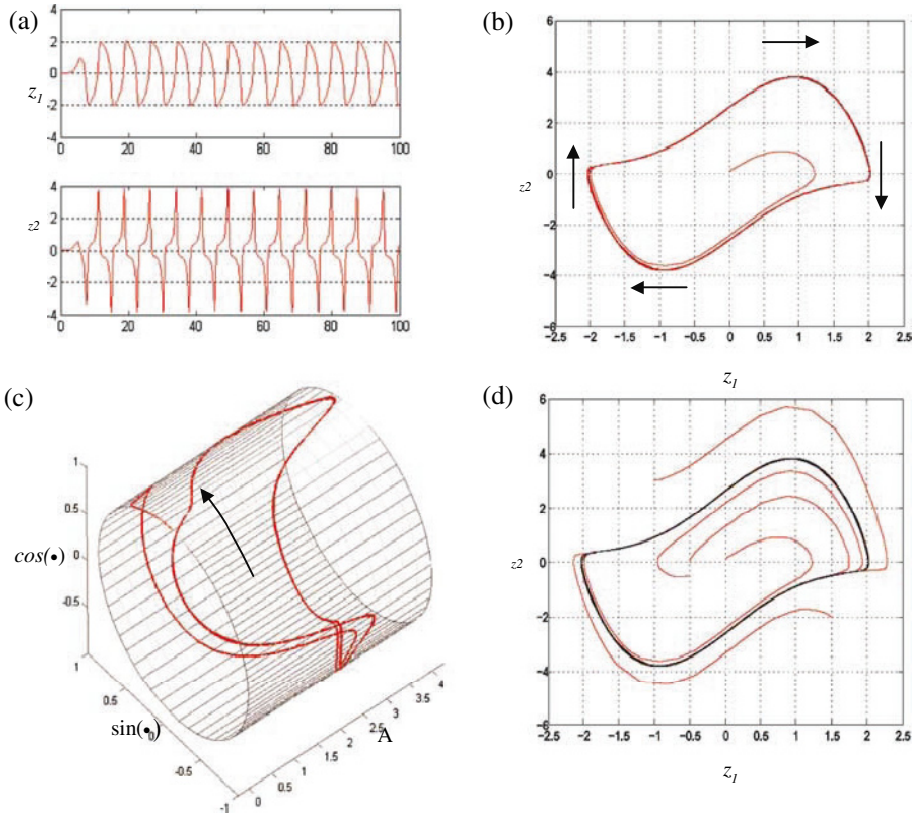


Fig. 4. Van der Pol oscillator for $\mu = 2$. (**a**) Time series. (**b**) Phase space portrait representation of a single orbit and its approach toward a limit cycle attractor. Arrows show representative vector field. (**c**) Representation of the same orbit on a cylindrical manifold spanned by the polar coordinates $A(t)$ and $\theta(t)$. (**d**) A set of distinct initial conditions flowing toward the limit cycle attractor (black)

two such points will stay forever close if on or near a limit cycle attractor (the largest characteristic exponent is zero). Two such points in the vicinity of a fixed point attractor will invariably get closer (the largest characteristic exponent is negative). A chaotic attractor is also ***topologically mixing*** – i.e. any given open set covering any region of the attractor will eventually overlap with any other region. The unceasing divergence of nearby orbits and the eventual mixing of regions combine to enable a chaotic attractor to be both unstable and bounded. We revisit chaos in the setting of specific neuronal models in Sect. 3.

An attractor's ***basin of attraction*** is the set of all initial conditions which have the attractor as their future state. In the present case, the basin of attraction for the loop is the entire plane, except for the origin. The ***inset*** of an attractor is that part of the basin of attraction with the strongest (principle)

direction of attraction. A ***repellor*** is an invariant set that is the past state of a large set of points, its ***basin of repulsion***. The ***outset*** of a repellor is the subspace of this basin which diverges most quickly. In the present case, the origin is a repellor and its basin of repulsion is all the points within the loop.

If time was reversed in (20), the origin would become an attractor and points within the loop would be its basin of attractions. Points outside the limit cycle would diverge towards infinity. Hence the loop would be an example of a ***basin boundary*** or ***seperatrix***. Basin boundaries can be repellers (as in the case here) or saddles which have an inset and an outset. A trivial example of a saddle is the origin in the system,

$$\frac{dz_1}{dt} = a_1 z_1 + b_1 \qquad \frac{dz_2}{dt} = a_2 z_2 + b_2 \tag{22}$$

where, for $a_1 = 1,\ a_2 = -1$ and $b_1 = b_2 = 0$ the z_1-axis is the inset and the z_2-axis is the outset (Fig. 5).

Occasionally, the outset from a repellor becomes the inset of an attractor. Such an entity, linking two fixed points, is called a ***heterocline***. For a saddle, it is possible that the outset becomes the inset (due to curvature away from the saddle point). If so, it is called a ***homocline***.

Two final concepts are required before we move onto more complex matters. An attractor possesses ***structural stability*** if it is insensitive to a small change in the nature of the vector field, corresponding to a small change in the evolution equation. "Insensitivity" here denotes that there exists a smooth mapping between the perturbed attractor and the original attractor. When such a mapping exists we say the two (original and perturbed) attractors are ***topologically conjugate***. The saddle point in Fig. 5 is structurally stable following changes in either the parameters a (under a stretching and/or contraction) and/or the parameters b (under a translation). Similarly the Van der Pol attractor in Fig. 4 is structurally stable since small changes to any of the parameters results in another (topologically conjugate) limit cycle attractor.

The ***nullclines*** of a dynamical system are the curves in phase space, for which one derivative in the evolution equation is equal to zero, and hence correspond to the regimes in phase space with zero flow in a particular direction.

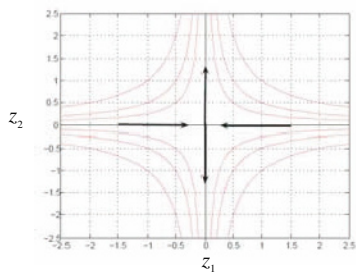


Fig. 5. Saddle point and orbits of (22)

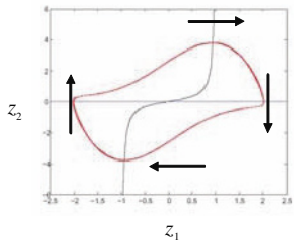


Fig. 6. Nullclines of the Van der Pol system. Arrows show direction of the vector field across the nullclines. The fixed point lies at the intersection of the two mullclines (in blue and black). The trajectory is shown in red

The two nullclines to the Van der Pol equation (20) are depicted in Fig. 6. The blue line shows the nullcline for zero flow in the z_1 direction and the black shows the curve for z_2. The limit cycle trajectory satisfies these conditions as it crosses the respective curves – that is, the $\mathrm{d}z_1/\mathrm{d}t = 0$ when the attractor (red curve) crosses the blue nullcline. By definition, any crossing of two nullclines corresponds to the existence and location of a fixed point since $\mathrm{d}z_1/\mathrm{d}t = \mathrm{d}z_2/\mathrm{d}t = 0$.

The nullclines form the "skeleton" of the phase space and, as we explore below, their intersections are vital to the existence and nature of most attractors, not just fixed points.

2.2 Bifurcations and Complex Dynamics

The preceding discussion captures the nature of phase space dynamics and its relationship to the evolution equations for a given set of parameter values (i.e. when the vector field is kept constant). An intriguing and important field of study concerns what happens to the attractors and basin boundaries following a change to the system's parameters and hence to the vector field. From above it follows that if all the attractors are structurally stable, then the effect of such a change can be considered trivial since the dynamics will remain qualitatively similar (and typically also quantitatively similar). However, in the case when this is not so, sudden and dramatic changes in the dynamics, denoted ***bifurcations***, occur. Examples abound in neuroscience, such as the generation of an action potential, the onset of bursting (Izhikevich 2005) and even the onset (Robinson et al. 2002, Lopes da Silva et al. 2003, Breakspear et al. 2006) and temporal progression (Rodriguez et al. 2006) of an epileptic seizure.

An important means of understanding the nature of a system's bifurcations is through the study of its ***bifurcation diagram***. This is produced by smoothly varying one parameter over some range of interest whilst keeping all other parameters fixed. Hence the vector field is smoothly changed in one dimension of parameter space. At each parameter value, the system is integrated and, after passage of an initial transient – allowing for the system to

evolve towards its attractor(s) – the asymptotic time series is captured. From this time series, the values of all local minima and maxima are stored. For a fixed point there will exist only one such value. For a simple (period-1) limit cycle there will exist two such points and for a period-2 oscillator, four such points – two maxima and two minima. For a chaotic oscillator, such points will be distributed ***densely*** ("almost everywhere") over one of more segments. The bifurcation diagram is the plot of these local maxima and minima against the respective parameter value. Figure 7 shows the bifurcation diagram of the logistic (4).

It is crucial to note that in most nonlinear systems, two or more attractors may co-exist for some parameter values, facilitating ***bistability*** or even ***multistability***. Each attractor will have basins, each separated by basin boundaries. In such cases, it is important that all such attractors are located when plotting a bifurcation diagram.

Bifurcations can be divided into local and global, as outlined below. Before doing so, it is important to introduce a second notion of stability. Structural stability concerns the robustness of invariant sets – attractors, repellors, saddles - to changes in the underlying vector field. In contrast ***asymptotic stability*** deals with the situation where the instantaneous state of the system is perturbed through addition of a small transient noise term (but the vector field is kept constant). An attractor is called asymptotically stable whenever the system returns towards the attractor following any such (small)

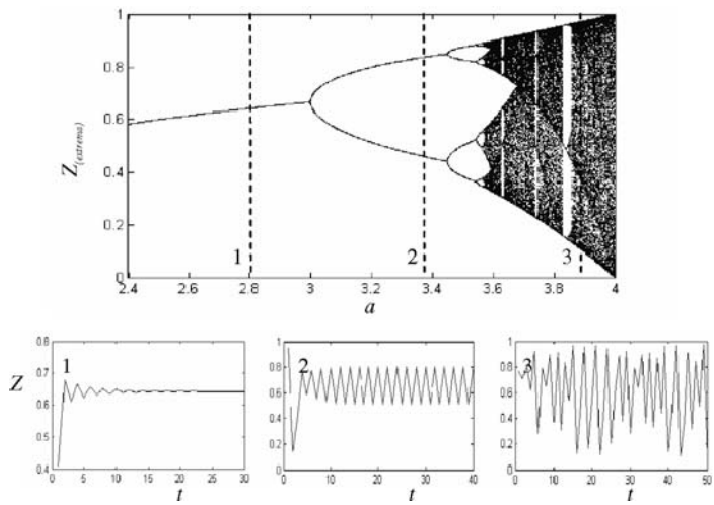


Fig. 7. Bifurcation diagram of the logistic equation (**4**). Top panel shows the local minima and maxima of the asymptotic time series against the parameter a. Lower three panels show representative time series (including the initial transient) with (**1**) fixed point, (**2**) limit cycle and (**3**) chaotic attractors. Note the "periodic windows" within the chaotic regime of the bifurcation diagram

noisy perturbation. A ***local bifurcation*** occurs whenever an attractor loses *asymptotic* stability whereas a ***global bifurcation*** corresponds to the loss of *structural* stability. These are also called *subtle* and *catastrophic* bifurcations (Abraham & Shaw 1988) because in the latter case the impact on the dynamics is typically more immediately discernable. We now explore such bifurcations in further detail.

Local Bifurcations

Local bifurcations concern the asymptotic stability of fixed point and other attractors. Consider the system governed by,

$$\frac{d\mathbf{Z}}{dt} = \mathbf{A}\mathbf{Z} + \mathbf{B}, \tag{23}$$

where $\mathbf{A}$ is a matrix and $\mathbf{B}$ a vector. This is the matrix form of (22). Solutions in the case where $\mathbf{B}$ is zero are of the form

$$\mathbf{Z}(t) = \mathbf{Z}(0)\, e^{t\mathbf{A}}. \tag{24}$$

Hence the origin is a fixed point and the eigenvalues Λ of $\mathbf{A}$ determine the nature of the neighboring flow. Solutions in the case $\mathbf{B} \neq 0$ are essentially the same after a suitable translation of the axes. The eigenvalues $\Lambda = \{\lambda_1,\ \lambda_2\}$ determine five possible types of fixed point systems (Fig. 8).

Figure 8(a) shows a typical flow when both eigenvalues are real and either both positive or both negative. Orbits diverge from ($\lambda_1 > 0,\ \lambda_2 > 0$) or converge to ($\lambda_1 < 0,\ \lambda_2 < 0$) the origin. In the former case, the fixed point is called a ***source*** and in the latter, a ***sink*** or ***node***. We have already met the case (Fig. 8b) where the eigenvalues are real and opposite in sign ($\lambda_1 > 0,\ \lambda_2 < 0$) for the ***saddle point*** discussed above with regards to basin boundaries. When the eigenvalues are complex, they occur as complex conjugate pairs. The imaginary component endows the time series with an oscillatory component evident as spiraling orbits (Fig. 8c). When the real part of each eigenvalue is negative, these oscillations are damped and the fixed point is a ***spiral inset***. Otherwise it is a ***spiral outset***.

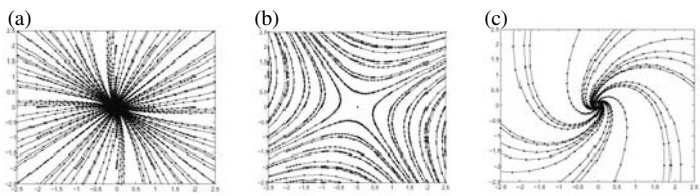


Fig. 8. Orbits for fixed points of the linear system (26). (**a**) Source or sink, (**b**) saddle, and (**c**) Spiral inset or outset

Whilst (26) is a simple linear system, the ***Hartman-Grobman theorem*** states that, for a very general class[5] of nonlinear systems $\mathbf{F}_a$, the flow within the neighborhood of a fixed point can be approximated by a suitable linear system with the form of (23). Hence these fixed points – and their stability – play an important role in many dynamical systems.

Note that the eigenvalues of $\mathbf{A}$ determine the divergence or convergence of nearby orbits. These are hence the "characteristic exponents" referred to in Sect. 2.1. In the setting of fixed points these are simply referred to as the eigenvalues of $\mathbf{A}$. They are often called ***Floquet exponents*** in the vicinity of a limit cycle and ***Lyapunov exponents*** for a chaotic attractor. Following Eckmann and Ruelle (1984), we will simply refer to them as characteristic exponents, whatever the nature of the invariant set to which they refer.

Local bifurcations hence deal with the zero crossings of the characteristic exponents of attractors. The underlying set (typically) remains invariant, but loses its asymptotic stability. Just as a zero crossing can transform a fixed point from an attracting node into a saddle, the same also applies for both limit cycles and chaotic attractors. We now briefly discuss some of the canonical local bifurcations. In sect. 3 we will see how they relate to fundamental neuronal events such as firing and bursting.

Canonical Local Bifurcations

Pitchfork bifurcations occur when a single fixed point changes its (asymptotic) stability whilst also splitting off extra fixed points. In a ***supercritical pitchfork bifurcation*** a single stable fixed point attractor loses its stability as a parameter crosses its threshold and two new stable fixed points appear (Fig. 9a). The evolution equation,

$$\frac{dz}{dt} = z\left(a - z^2\right), \tag{25}$$

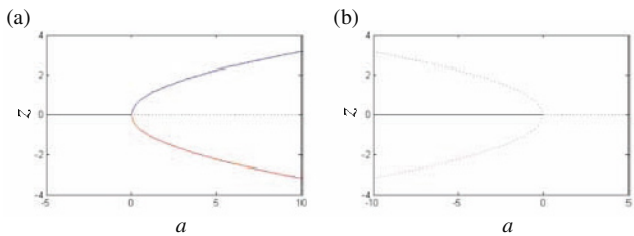


Fig. 9. Pitchfork bifurcation diagram (**a**) Supercriticial, and (**b**) subcritical. Solid line denotes fixed point attractor. Dashed lines denote fixed point repellors

[5] As long as the derivative of $\mathbf{F}_a$ at the fixed point is not zero – i.e. the fixed point is *hyperbolic*.

yields this type of bifurcation at $a = 0$. Note that for $a < 0$, we have $dz/dt < 0$ when $z > 0$ and $dz/dt > 0$ when $z < 0$. Hence all initial conditions lead to the fixed point $z = 0$. Similar calculations show that when A crosses zero ($a > 0$) the origin becomes a source and fixed point attractors exist as $\pm\sqrt{a}$.

On the other hand the equation,

$$\frac{dz}{dt} = z\left(a + z^2\right), \tag{26}$$

yields a ***subcritical pitchfork bifurcation***. In this case, the fixed point attractors at $z = 0$ also loses its stability as A crosses zero from below. However, two fixed point repellors exist at $\pm\sqrt{-a}$ when $a < 0$ (Fig. 9b). Looking at the situation alternatively, one could say that the fixed point attractor at $z = 0$ loses its stability when two fixed point repellors collide with it at $a = 0$. However, in both cases, the fixed point remains an invariant of the system (i.e. $dz/dt = 0$) for all a.

In a ***transcritical bifurcation***, there are two equilibrium points which collide and exchange their stability at the bifurcation point. For example, the evolution equation,

$$\frac{dz}{dt} = z\left(a - z\right), \tag{27}$$

has two equilibrium points, the origin $x = 0$ and $x = a$. When $a < 0$, the origin is an attractor but becomes a repellor as a crosses zero (Fig 10).

A ***Hopf bifurcation*** (Fig. 11) is much like a pitchfork bifurcation with the exception that it involves a limit cycle attractor. Hopf bifurcations play an important role in neuronal models as they describe the onset of both subthreshold membrane oscillations and cell firing. Consider the equation,

$$\frac{dz}{dt} = z\left(a + b\left|z\right|^2\right), \tag{28}$$

where a is the bifurcation parameter and both z and b are complex numbers. When the real part of b is negative then the system exhibits a ***supercritical Hopf bifurcation*** (Fig. 11a,c–f). For $a < 0$ there exists a single stable fixed point attractor (a spiral inset). When $a > 0$ this fixed point is an unstable spiral outset and there also exists a stable limit cycle.

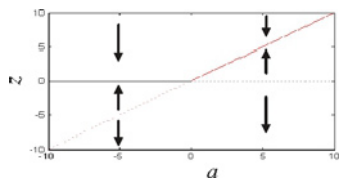


Fig. 10. Transcritical bifurcation. Solid line denotes fixed point attractor. Dashed lines denote fixed point repellors. Arrows show representative vector field

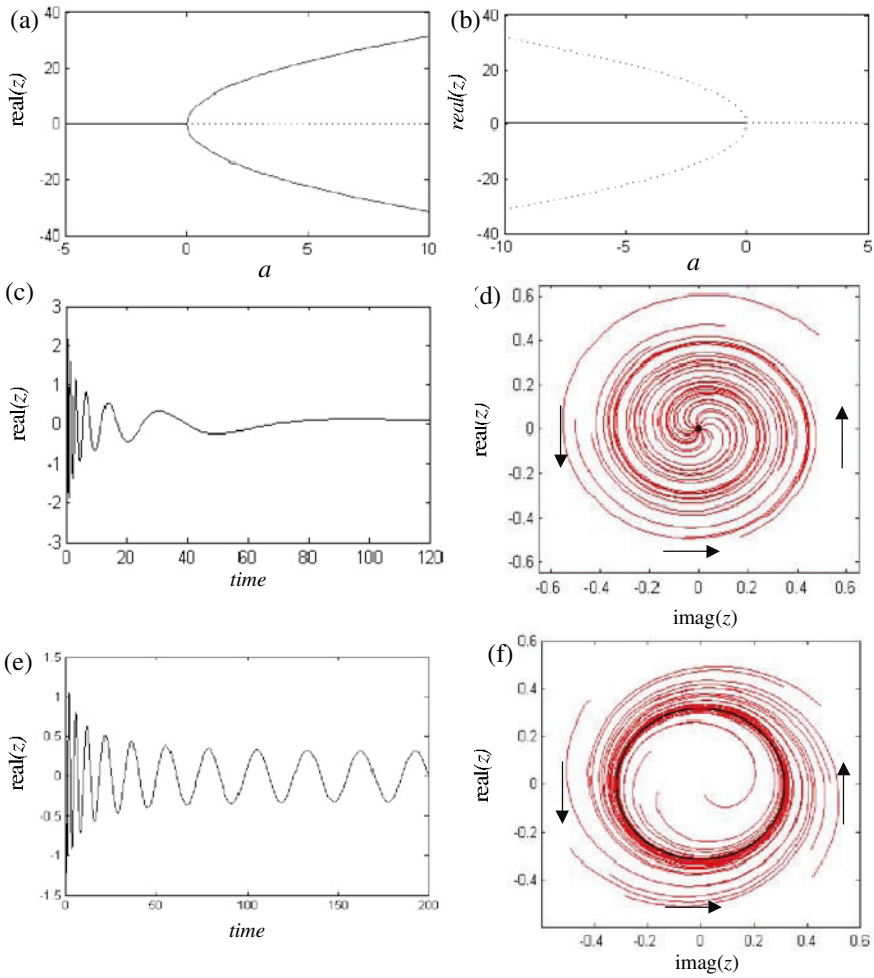


Fig. 11. Hopf bifurcation (**a**) Supercriticial, and (**b**) subcritical. Black line denotes fixed point attractor (solid) and repellor (dashed). Blue lines denote the maxima and minima of the limit cycle attractor (panel a; solid) and repellor (panel b; dashed). Time series (**c**) and phase space portrait (**d**) of the fixed point attractor (**e**) in the supercritical system (i.e. when $a < 0$ and $\text{real}(b) < 0$). Time series and phase space portrait (**f**) of the limit cycle attractor (**black**) in the supercritical system (i.e. when $a > 0$ and $\text{real}(b) < 0$). Red orbits show transients

Conversely, when the real part of b is positive then the system exhibits a ***subcritical Hopf bifurcation*** (Fig. 11b). For $a < 0$ there exists a single stable fixed point attractor (a spiral inset) and an unstable periodic orbit. Hence the phase space is partitioned: those initial conditions within the periodic orbit spiral in towards the fixed point; those initial conditions outside of the limit cycle diverge towards infinity. When $a > 0$, there exists an unstable

fixed point (a spiral outset). Hence all initial conditions (except $z = 0$) diverge towards infinity.

A subcritical Hopf bifurcation often occurs in the context of bistability, when there co-exists a large-amplitude limit cycle attractor. Above the bifurcation point $a > 0$, orbits diverge outwards from the fixed point repellor to this attractor. Below the bifurcation point $a < 0$, the limit cycle repellor (dashed blue curve in Fig. 11b) separates the basin boundaries of the fixed point attractor at the origin and the large-amplitude limit cycle attractor.

In summary, Hopf bifurcations are of very high importance for an understanding of neural activity as they explain the onset and nature of oscillatory behaviour. Supercritical Hopf bifurcations lead to the appearance of small amplitude periodic oscillations. Subcritical Hopf bifurcations result immediately in a large amplitude limit cycle.

Period-doubling bifurcations typically occur as a sequence of events subsequent to a Hopf bifurcation, following a further increase in the bifurcation parameter. The main panel in Fig. 7 shows period-doubling bifurcations in the logistic map between parameter values $a \sim 3.0$ and the onset of chaos at $a \sim 3.5$. The first period-doubling bifurcation (i.e. from a simple periodic oscillation to a period-2 oscillation) corresponds to the "excitation" of the limit cycle attractor into an invariant torus around which the attractor winds. Subsequent period-doubling bifurcations increase the number of times the attractor twists around the short axis of the torus every time it makes one complete revolution around the long axis.

A ***fold*** or ***saddle-node*** is an interesting and illustrative bifurcation. It occurs when a stable and unstable fixed point collide (see Fig. 12). Consider the equation,

$$\frac{dz}{dt} = a + z^2. \tag{29}$$

For $a < 0$ there are two fixed points, an attractor at $-\sqrt{-a}$ and a repellor at $+\sqrt{-a}$. As a approaches zero from below, these two fixed points hence approach each other and collide at $a = 0$. At this point, the fixed point, at $z = 0$, is attracting for $z < 0$ and repelling for $z > 0$. Hence it is neither an attractor nor a repellor, but rather a special ("non-hyperbolic") fixed point called a "saddle-node". The saddle node is classified as a local bifurcation because the two fixed points lose their asymptotic stability when they collide at $a = 0$. However, at $a > 0$ there is no fixed point – i.e. there is also a loss of structural stability. Hence it also meets the criteria for a global bifurcation.

A recently described phenomena is the ***blowout bifurcation*** (Ott & Sommerer 1994). Suppose we have two (n-dimensional) neural systems $z_1(t)$ and $z_2(t)$ evolving according to (11), each with chaotic dynamics. In the instance where there is no coupling ($c = 0$), the system as a whole explores the full phase space $\mathrm{R}^n \times \mathrm{R}^n$. When the system is strongly coupled ($c > c_k$ for some threshold coupling strength c_k), then the two systems will synchronize (Fujisaka & Yamada 1983, Pecora & Carroll 1990). In this case, the dynamics are confined to the 'hyper-diagonal' – that is, the space $z_1(t) = z_2(t)$ of half

the dimension of the full phase space. If the coupling strength c falls below c_k the dynamics of the two-cell system "blowout" from the low dimensional state of synchronization, into the full state space, as shown in Fig. 13. Alternatively, it can be said that when $c > c_k$ the state of synchronization has asymptotic stability. When c crosses c_k from above, the system loses stability in the direction transverse to the hyper-diagonal (Ashwin et al. 1996). Looking at the blowout bifurcation as c crosses c_k from below (i.e. as the coupling strength is increased) we see that – through the process of synchronization - the dynamics of the system collapse onto a (relatively) low dimensional manifold. That is, synchronization constrains the number of degrees of freedom of a spatially distributed dynamical system.

Global Bifurcations

Whereas local bifurcations deal with the loss of asymptotic stability of fixed points - and are hence concerned with the dynamics in local neighborhoods of attractors - global bifurcations can only be understood by studying the properties of the vector field outside of such neighborhoods. They occur when an attractor loses structural stability. Their nature depends upon the "skeleton" of the phase space – the nullclines, homoclines and heteroclines.

As stated above, a saddle-node bifurcation does have the properties of both a local and a global bifurcation in that there is no fixed point for $a > 0$. Looking at Fig. 12, we see that there are two fixed points when $a < 0$, an attractor at $-\sqrt{-a}$ and a repellor at $+\sqrt{-a}$. The latter forms the boundary of the basin for the former. Hence the bifurcation occurs when the attractor collides with its basin boundary. This collision illustrates the principles of more complex global bifurcations. An example, when a chaotic attractor collides with its fractal basin boundary was termed a ***crisis bifurcation*** when it was first described (Celso et al. 1982). More specifically, a ***boundary crisis*** occurs when the collision is between an attractor and a boundary, whereas an ***interior crisis*** results from the collision between an attractor and a saddle. The former results in the sudden loss of an attractor and its basin, whereas the latter typically leads to the sudden increase in the extent of the attractor.

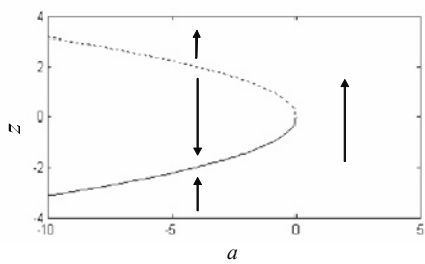


Fig. 12. Saddle-node bifurcation. Solid line denotes fixed point attractor. Dashed line denote fixed point repellors. Arrows show representative vector field

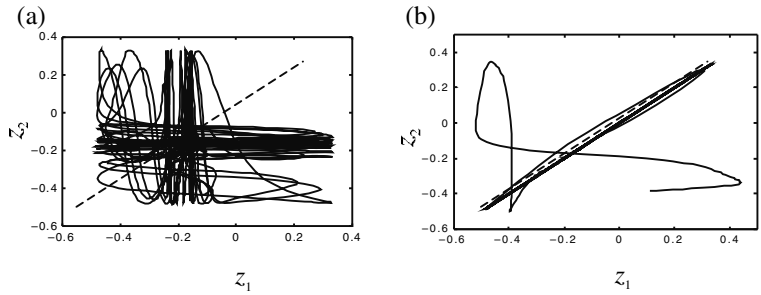


Fig. 13. Blowout bifurcation in system of two coupled logistic maps. (**a**) When $c < c_k$ the system explores R^2. (**b**) When $c > c_k$ the system contracts onto the diagonal after a brief transient. Arrow points to initial condition. A "blowout bifurcation corresponds to a transition from (**b**) to (**a**). Adapted from Breakspear (2004)

Other global bifurcations involve intersections of homoclinic or heteroclinic orbits either with themselves, or with fixed point attractors. A fascinating example involves the birth of a limit cycle out of a saddle-node bifurcation on a homocline! This obscure-sounding event actually lies at the heart of neuronal firing and we discuss it in more depth in the next section.

3 A Taxonomy of Neuronal Models

Neurons are traditionally seen as the building blocks of the brain. It hence makes sense to gain some insight into their dynamics – and functional interactions – at the microscopic scale at which they reside before moving into the larger scales, which we do in Sect. 4.

The "foundation stone" of microscopic models are the conductance-based Hodgkin-Huxley model and its derivatives. A full description of these is provided by a number of authors (e.g. Izhikevich 2005, Gerstner & Kistler 2002, Guevara 2003). Our objective here will be to quickly move from the full model to a two dimensional approximation and then explicate the onset of neuronal firing as a dynamical bifurcation.

3.1 The McCulloch-Pitts System

Before we do this, for the sake of theoretical and historical completeness, we briefly discuss the McCulloch-Pitts model (1943),

$$z(x_i, t+1) = S\left(\sum_j h_{ij} z(x_j, t) - \varepsilon_i\right), \tag{30}$$

where h_{ij} is the connectivity matrix, ε the "threshold" of neuron i and S is the step function. Neural inputs to a given unit are summed and then

converted into a binary output if they exceed the threshold ε_i. The resulting output is iteratively fed back into the network. Hence the McCulloch-Pitts model is discrete in both space and time, and as such is an example of a coupled difference map (5). Considered together with the use of the step function as representing neural "activation", this model is perhaps as abstract as possible. Nonetheless, McCulloch and Pitts proved that the system was capable of remarkably general computational feats. However, it is probably fair to say that this model finds its place more appropriately in the lineage of artificial neural networks than in the understanding of the dynamics of biological systems. Hence, McCulloch-Pitts systems form the basis for the two-layer "perceptrons" of Rosenblatt (1958) and the symmetric Hopfield (1982) networks. These extend the complexity and computational properties of McCulloch-Pitts systems to permit object categorization, content-addressable memory (i.e. the system correctly yields an entire memory from any subpart of sufficient size) and learning. For example, Sejnowski and Rosenberg (1987) showed that such systems, if constructed with three interconnected layers, are able to learn language pronunciation.

An overview of related advances is provided by Ermentrout (1998). For a fascinating history of this model and the life of Walter Pitts, see Smalheiser (2000).

3.2 Biophysical Models of the Neuron: The Hodgkin-Huxley Model

Whereas the McCulloch-Pitts system was constructed to embody only the very general network properties of neural systems and to directly address computational issues, the Hodgkin Huxley model aims to incorporate the principal neurobiological properties of a neuron in order to understand phenomena such as the action potential. Computational properties of these neurons are then investigated.

The paper of Hodgkin and Huxley (1952) is remarkable in that it casts detailed empirical investigations of the physiological properties of the squid axon into a dynamical systems framework. The Hodgkin-Huxley model is a set of conductance-based coupled ordinary differential equations[6] of the form of equation (8), incorporating sodium (Na), potassium (K) and chloride ion flows through their respective channels. Chloride channel conductances are static (not voltage dependent) and hence referred to as leaky (L). Hence we have,

$$C\frac{dV(t)}{dt} = g_{Na} f_{Na}(V(t)) \times (V(t) - V_{Na}) + g_K f_K(V(t)) \times (V(t) - V_K) + g_L \times (V(t) - V_L) + I, \tag{31}$$

where $c = 1\ \mu\mathrm{F/cm}^2$ is the membrane capacitance, I is an applied transmembrane current and V_{ion} are the respective Nernst potentials. The

[6] Here we depart slightly from the traditional nomenclature in order to simplify the mathematical description of the model.

coefficients g_{ion} are the maximum ion flows in the case where all the channels of that ion species are open. The Na and K ion flows reflect the state of "activation" channels, which open as membrane voltage increases and "inactivation" channels, which close. These are given by,

$$f_{Na}(V) = m(V)^M h(V)^H$$
$$f_K(V) = n(V)^N \tag{32}$$

where m and n are activation channels for Na and K, and h is the single inactivation channel for Na. The exponents are determined by the number of such classes of channel $M = 3$, $H = 1$ and $N = 4$. Hence (31) reflects the combined flow of all ion species as they are "pushed through" open channels according to the gradient between the membrane and Nernst potentials.

The kinetics of activation and inactivation channels are determined by differential equations of the form,

$$\frac{dm(V)}{dt} = \frac{(m_\infty(V) - m(V))}{\tau_m(V)} \tag{33}$$

where $m_\infty(V)$ is the fraction of channels open if the voltage is kept constant and $\tau_m(V)$ is a rate constant. These are determined empirically. These equations embody the exponential relaxation of channels towards their (voltage-dependent) steady states $m_\infty(V)$ consequent to a transient change in membrane potential. The kinetics of h and n are of the same form, although their rate constants τ are obviously distinct. The form of $m_\infty(V)$ – the steady state configurations of ion channel populations as a function of membrane potentials – is sigmoid shaped of the form,

$$m_\infty(V) = \frac{m_{\max}}{1 + e^{(V_m - V)/\sigma}}. \tag{34}$$

where V_m is the threshold potential for the ion channel and σ introduces the variance of this threshold. Figure 14 summarizes membrane dynamics in the Hodgkin-Huxley model.

The Hodgkin-Huxley model is a conductance approach to the dynamics of neural activity, reflecting ion flows through voltage- and time-dependent transmembrane channels. It represents a beautiful juncture of empirical and mathematical analysis. It not only offers an explanation of neural firing, but it quantitatively captures the complex shape of a neural depolarization.

3.3 Dimension Reductions of the Hodgkin-Huxley Model

The Hodgkin-Huxley model is able to explain the chief properties and many of the nuances of neuronal depolarization, including the threshold effect and the post-depolarization refractory period with quantitative accuracy. However, much of the qualitative (and some of the quantitative) behaviour can be

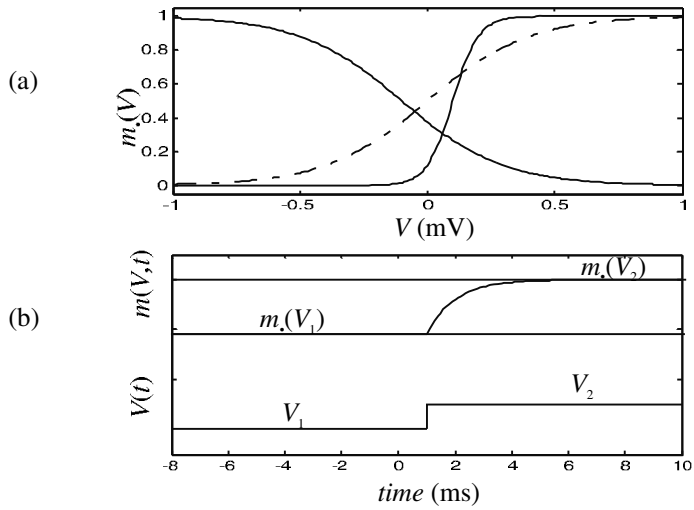


Fig. 14. Transmembrane currents in the Hodgkin-Huxley model. (**a**) Three examples of the sigmoid relationship between transmembrane potential V and steady state conductances $m_\cdot$. Solid and dotdashed lines denote "activation" channels (such as m and n) whereas the dashed line denotes an inactivation channel (such as h) (**b**) Exponential "relaxation" of transmembrane conductance according to (33), following a discrete change in the transmembrane potential (lower line). Conductance m 'relaxes' from $m_\cdot(V_1)$ to $m_\cdot(V_2)$

captured by greatly reduced approximations. Amongst other things, first pass approximations ignore the voltage-dependent nature of τ_m and make further simplifications, but are still able to capture many of the important dynamics, such as neural depolarization. We now describe these, following the basic approach of Izhikevich (2005).

Morris-Lecar and Related 'Planar' Simplifications

An essential ingredient of a neural firing is a fast depolarizing current such as Na^+ – which is turned on subsequent to a synaptic current - and a slow repolarizing current such as K^+ - which restores the resting membrane potential. These in turn are facilitated by the existence of slow and fast ion channels of the respective species, $\tau_m(V) << \tau_n(V)$. The depolarizing current represents positive feedback (i.e. is self promoting) and, if a threshold is reached before a sufficient number of slower K^+ channels are open, the cell depolarizes. By contrast, the Na^+ inactivation channel plays less of a "brute force" role and can be ignored. The requirement of a "fast" depolarizing current and a slow repolarizing current can be met in a two dimensional ("planar") system,

$$\frac{dV}{dt} = g_{Na} m_\infty (V) \times (V - V_{Na}) + g_K n(V) \times (V - V_K) + g_L \times (V - V_L) + I, \tag{35}$$

where the dynamics of the slow repolarizing K^+ is given by

$$\frac{dn}{dt} = \frac{(n_\infty - n)}{\tau_n} \tag{36}$$

and the steady state currents given by,

$$n_\infty(V) = \frac{n_{\max}}{1 + e^{(V_n - V)/\sigma}}, \qquad \text{and} \qquad m_\infty(V) = \frac{m_{\max}}{1 + e^{(V_m - V)/\sigma}},$$

In other words, fast sodium channels *instantaneously* assume their steady state values following a change in membrane potential, hence adapting in a step-wise manner to a step-like change in membrane potential. Hence there is no differential equation for the Na^+ activation channels, m. This is exactly the form of the Morris-Lecar model, with the exception of a substitution of Na^+ currents with Ca^{++}.

The system (35)–(36) is known as planar, as its phase space is the two-dimensional plane spanned by V (the abscissa) and n (the ordinate). To understand the dynamics we calculate the nullclines for the dynamical variables V and n. The V-nullcline, obtained by substituting $\mathrm{d}V/\mathrm{d}t = 0$ into (35) is,

$$n = \frac{I - g_{Na} m_\infty(V) \times (V - V_{Na}) - g_L \times (V - V_L)}{g_K \times (V - V_K)}. \tag{37}$$

Similarly, the n-nullcline, obtained by setting $\mathrm{d}n/\mathrm{d}t = 0$ is,

$$n = \frac{n_{\max}}{1 + e^{(V_n - V)/\sigma}}, \tag{38}$$

These nullclines for the parameter values given in Table 1 and synaptic current $I = 0$ are plotted in Fig. 15. We see that there are three nullclines crossings corresponding to three fixed points, $\{-66, 0\}$, $\{-56, 0\}$ and $\{-25, 0.5\}$. Stability analysis shows that these fixed points are a stable focus, saddle point and spiral outset respectively (Izhikevich 2005). Hence the first fixed point first represents the only stable (steady state) solution.

Figure 15(b) shows two heteroclines – that is, outsets of the saddle point that become insets of the stable node. A long heterocline (magenta) traverses the nullclines before reaching the node whereas the shorter one (yellow) is able to track in parallel to the n-nullcline directly between the fixed points.

Table 1. Parameter values for the planar system (37)–(38) and figures 15–19

Capacitance,	$C = 1$;
Synaptic current (default),	$I = 0$;
Leaky channels:	$V_L = -80;\ g_L = 8;$
Sodium channels:	$V_{Na} = 60;\ g_{Na} = 20; V_m = -20; \sigma_m = 15; m_{max} = 1;$ $\tau_m = 1;$
Potassium channels:	
"high threshold"	$V_K = -90;\ g_K = 10;\ V_n = -25;\ \sigma_n = 5;\ n_{max} = 1;\ \tau_n = 1;$
"low threshold"	$V_K = -78;\ g_K = 10;\ V_n = -45;\ \sigma_n = 5;\ n_{max} = 1;\ \tau_n = 1;$

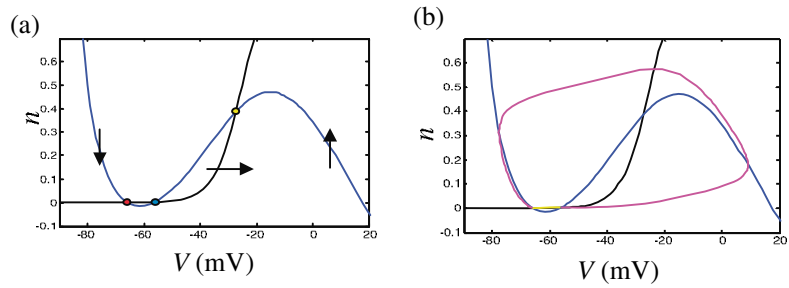


Fig. 15. Fixed points and nullclines of the planar system (37)–(38). The V-nullcline (39) is given in blue and the n-nullcline (eqn 40) in black. (**a**) Fixed points occur at the intersections of the nullclines: Stable node (red), saddle point (blue) and spiral outset (yellow). Arrows show representative vector field. (**b**) Long (magneta) and short (yellow) heteroclines of stable node and saddle

In Fig. 16 is shown representative orbits of this system. Three "subthreshold" (green) and three "suprathreshold" (red) orbits are shown. In the latter case, the neuron depolarizes before returning to its resting state. It should be noted that this threshold depends not only on the initial membrane potential V but also the initial K^+ membrane conductance. The separatrix between sub- and supra-threshold is constituted by the inset of the saddle point (not shown).

Whether the initial condition is sub- or supra-threshold, this system only has a single steady state solution in the current parameter regime. Hence, after at most one depolarization, it enters a quiescent state. Thereafter a discrete synaptic input, such as due to an excitatory post-synaptic potential (EPSP), will trigger a further discharge only if it is of sufficient strength to 'knock' the system over the inset of the saddle point. This will hence determine whether the resulting neural response is of the green or red waveform as in Fig. 16.

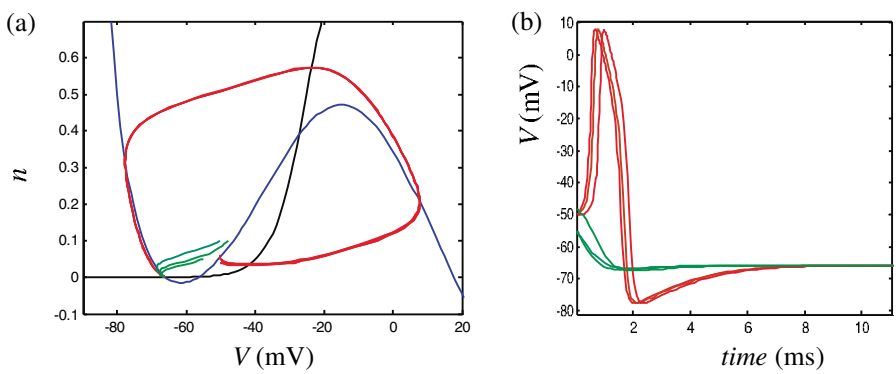


Fig. 16. (**a**) Representative sub- (green) and supra-threshold orbits (red) and (**b**) their temporal evolution

Bifurcations and Neuronal Firing in Planar Models

Saddle-node bifurcation

A further examination of the equation for the V-nullcline (37) shows that the synaptic current is a purely additive term. It hence acts to translate this nullcline in the vertical direction, with no influence on its shape and no influence on the n-nullcline. In Fig. 17, a close-up of the nullclines is shown for values of $I = 0$, 2, 4.51 and 6. As I is increased from 0 to 2 (dot-dashed), we see an upward shift of the V-nullcline so that the saddle and node fixed points are closer together in phase space. At $I = 4.5$ (dashed), the nullclines are tangent and the fixed points have hence collided. At $I = 6$ (dotted) there are no nullcline intersections: hence their collision has led to their mutual annihilation!

This is exactly the "saddle-node" bifurcation defined at Fig. 12. In the present setting, the synaptic input I functions as the bifurcation parameter. However, in addition to the structure of Fig. 12, an additional "global" feature of the phase space in the current system requires consideration. When the fixed points collide, the short heterocline is abolished, but the long heterocline[7] remains (Fig. 17b). Indeed even when $I > 4.51$ this orbit is still an invariant of the dynamics. However, with no fixed point along its domain, it is now a continuously looping limit cycle.

Figure 18 shows the limit cycle attractor (red) and its temporal dynamics for $I = 4.75$ (top row) and $I = 6$ (bottom row). Note that although the phase space portraits look similar, the frequency of the dynamics increases substantially with the increase in synaptic current.

This can be understood as a consequence of the bifurcation. Just after the bifurcation, although the nullclines do not intersect, the limit cycle must pass

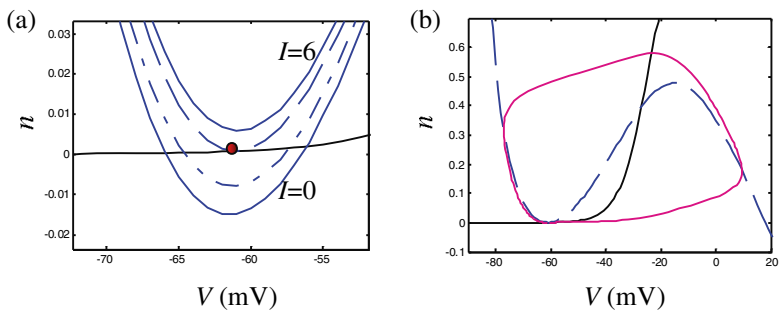


Fig. 17. Saddle-node bifurcation in the planar system. (**a**) Nullclines near fixed points for $I = 0$, 2, 4.51, 6. Red circle denotes "saddle-node" fixed point (**b**) Homoclinic orbit for the system when $I = 4.51$

[7] In fact, as there exists only a single fixed point, this orbit is more accurately now a "homocline".

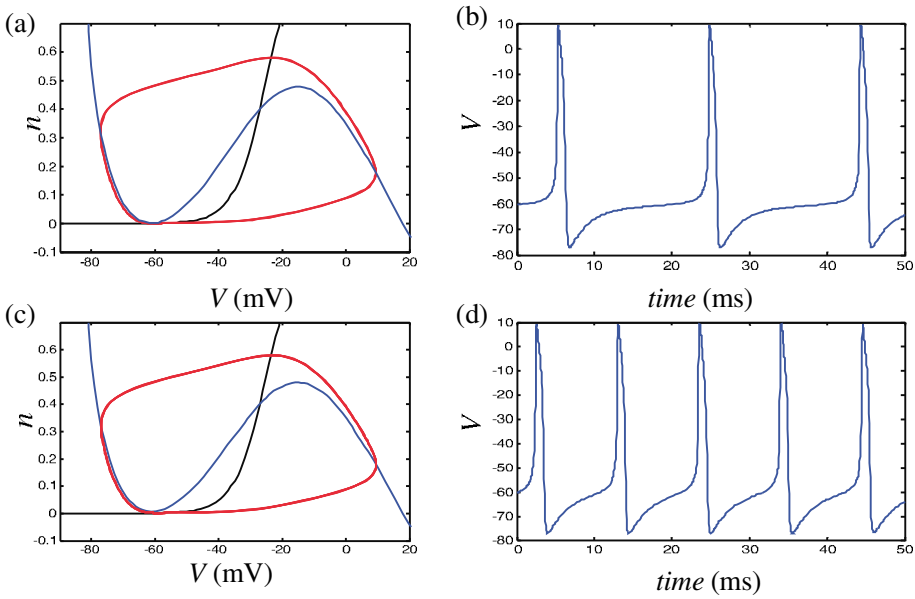


Fig. 18. Limit cycle dynamics for $I = 4.75$ (top row) and $I = 6$ (bottom row)

through a very narrow gap between them. The vector field in this gap bears the "memory" of the fixed points – namely it is very slow. Hence the orbits in this vicinity are near-stationary, as can be seen in the time domain. As I increases this influence diminishes and the frequency hence increases. Note that in both cases, however, there is virtually no change in the morphology of the depolarization, which is not related to this phenomenon.

Hopf bifurcation

Through a slight change in the parameters relating to the potassium channels, however, the transition from steady state (fixed point) to periodic (limit cycle) dynamics can occur through a different type of bifurcation. In the above scenario the potassium channels had values consistent with a "high threshold", namely the mean threshold potential of the K^+ potassium channels $V_n = -25\,\text{mV}$. Lowering V_n to $-45\,\text{mV}$ and changing the Nernst potential to $V_K = -78\,\text{mV}$ yields the phase space portraits and time series plotted in Fig. 19.

Firstly, there is only one interception of the nullclines for these parameter values, and hence only one fixed point. For $I < 19$ this is a spiral inset, hence yielding damped oscillations (panels a,b). For $I > 19$ the fixed point has undergone a (supercritical) Hopf bifurcation, hence yielding a small amplitude limit cycle, coinciding with sustained but subthreshold voltage oscillations. For $I \sim 26$, the amplitude of these oscillations grows smoothly but rapidly so that with $I = 27$ the system exhibits sustained suprathreshold oscillations. However, note that the damped, subthreshold and suprathreshold oscillations

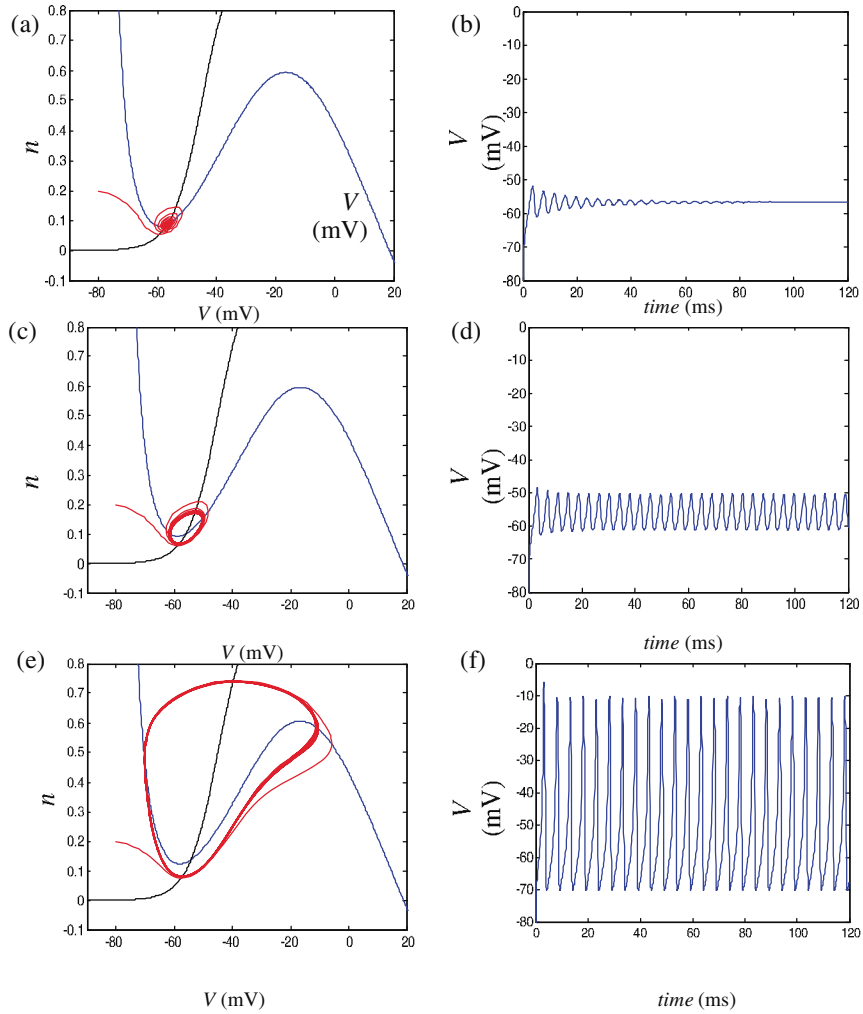


Fig. 19. Phase portrait and time series of the planar model in the "low K^+ threshold" case for $I = 18.5$ (top row), $I = 21$ (middle row) and $I = 27$ (bottom row)

all have approximately the same frequency. This contrasts with the saddle-node scenario.

We conclude with the two different bifurcation sets, Fig. 20, corresponding to distinct routes to sustained oscillations in this neuronal model. Panel (a) shows the saddle-node bifurcation, yielding the sudden onset of suprathreshold oscillations at $I \sim 4.5$ mA. Panel (b) depicts the Hopf bifurcation with the gradual onset of subthreshold oscillations at $I \sim 19$ mA, growing rapidly to suprathreshold with $I \sim 26$ mA.

In the presence of discrete synaptic inputs, the saddle-node system will generate an all-or-nothing depolarization – or chain of depolarizations – if

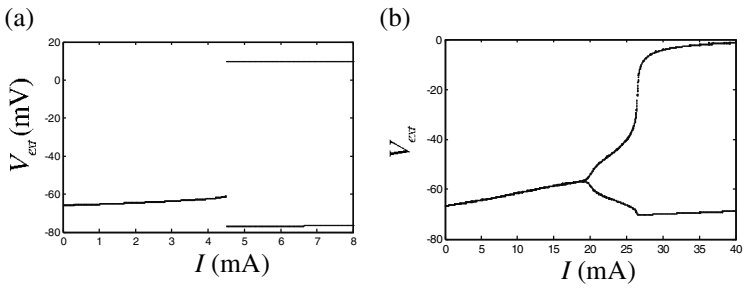


Fig. 20. Saddle-node (**a**) and (**b**) Hopf bifurcation diagrams for the planar neural system with high and low K^+ channel thresholds, respectively

the input is sufficiently large. The frequency of any such chain of discharges increases with the magnitude of the synaptic input. On the other hand, the Hopf route will generate either damped, sub-threshold oscillations or a chain of depolarizations, although the frequency of these will be more-or-less constant. In the presence of discrete synaptic inputs, the saddle-node system will generate an all-or-nothing depolarization – or chain of depolarizations – if the input is sufficiently large. The frequency of any such chain of discharges increases with the magnitude of the synaptic input. On the other hand, the Hopf route will generate either damped, sub-threshold oscillations or a chain of depolarizations, although the frequency of these will be more-or-less constant. As discussed in Izhikevich (2005) these two distinct neuronal responses to applied (or synaptic) currents were first observed empirically by Hodgkin in the 1940's. Specifically, he classified neurons that showed a frequency-dependence on the size of the synaptic current (i.e. Hopf-like responses) as *Type I neurons*. In particular, for small currents, these neurons begin to fire at very slow frequencies. In contrast, those neurons that start firing at relatively rapid rates following a supra-threshold input – and which show very little further increases in frequency – were classified as *Type II neurons*. The squid axon described by the original Hodgkin-Huxley model (1952) is a representative example of a neuron with type II behavior.

The FitzHugh-Nagumo Model

As we have seen above, the shape and intersections of the nullclines plays the determining role in the behavior and bifurcations of the dynamics. In fact, all that is required to reproduce the qualitative nature of the dynamics is the cubic-like shape of the V-nullcline and the presence of an n-nullcline with the appropriate intersections. Mathematically, these requirements can be met with the much simpler algebraic equations (FitzHugh 1961, Nagumo et al. 1962),

$$\frac{dx}{dt} = x\,(a - x)\,(x - 1) - y + I, \qquad \frac{dy}{dt} = bx - cy, \tag{39}$$

which have the simple nullclines,

$$y = x\,(a - x)\,(x - 1) + I, \qquad y = {}^{b}\!/_{c}\,x, \tag{40}$$

In Fig. 21 is illustrated a phase portrait and time series for this system following a super-critical Hopf bifurcation of the single fixed point. This system - and variations of it - are known as the FitzHugh-Nagumo model.

This system hence allows a closed-form analysis, with relatively simple algebraic forms, of the same qualitative phenomena as the planar model of Hodgkin-Huxley dynamics.

The Hindmarsh-Rose Model

The Hindmarsh-Rose model is the last in the "microscopic" domain for consideration. It continues the logic of the FitzHugh Nagumo model – namely that it captures the qualitative essence of neuronal firing through a simple algebraic form of the evolution equations (and hence of the nullclines). However, rather than further reducing the Hodgkin-Huxley model, the Hindmarsh-Rose (1984) model introduces an extra property. The system is given by,

$$\frac{dx}{dt} = y - ax^3 + by^2 - z + I, \tag{41}$$

$$\frac{dy}{dt} = c - dx^2 - y, \tag{42}$$

$$\frac{dz}{dt} = r\,[s\,(x - x_0) - z]\,, \tag{43}$$

When $r = 0$, the third variable plays no role and the system reduces to a variation of a FitzHugh Nagumo model – that is, a two dimensional spiking neuron with a simple algebraic form: An example is given in Fig. 22.

However, setting $r > 0$ but small has the effect of introducing the third variable into the dynamics. Notice that z only enters into the first two

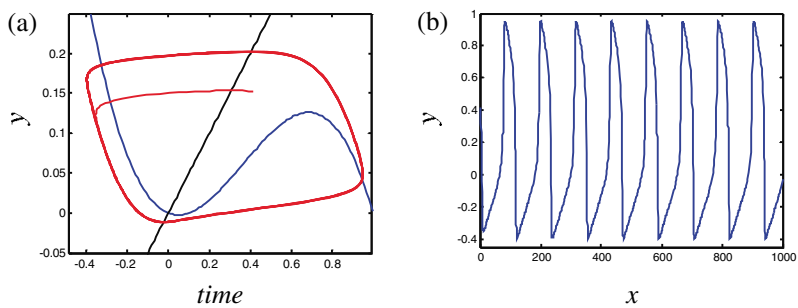


Fig. 21. Phase portrait and time series for the Fitz-Hugh-Nagumo model following a Hopf bifurcation. Parameters, b = 0.01; c = 0.02; a = −0.1; I = 0.1 as per Izhikevich (2005). Note the qualitative similarity to Fig. 19(e), (f)

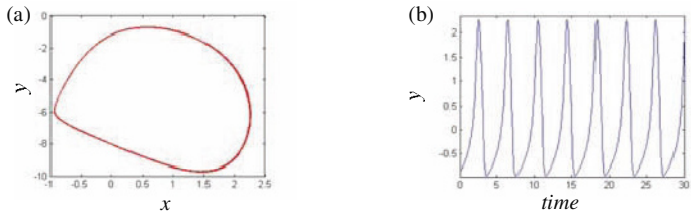


Fig. 22. Phase portrait and time series for the Hindmarsh-Rose model. Parameters, I = 3.2; a = 1.0; b = 3.0; c = 1.0; d = 5.0; s = 4.0; x0 = −1.60; r = 0

equations as an additive term – the same as I although negative in size. Also, setting r small has the effect of ensuring that z evolves on an intrinsically slower time scale than x and y. Together, these constructions have the effect of ensuring that z acts like a slowly varying synaptic current, albeit one which, due to the x term in (45), is also state dependent. Hence as z becomes more negative, it acts like the bifurcation parameter in the FitzHugh Nagumo model and precipitates – via a subcritical Hopf bifurcation - a run of depolarizations. However, due to the x term in (43), these depolarisations have the (relatively slow) effect of increasing z. Eventually the depolarizations are terminated as the reduced effective contribution of z to total synaptic current restabilizes the fixed point via a saddle node bifurcation. Hence the system, as shown in Fig. 23, exhibits a burst of spikes interspersed by quiescent phases. Indeed with $r = 0.006$, the system exhibits this pattern in a chaotic fashion.

Note that, as discussed, the fast spikes are far more evident in the dynamics of x whereas the dynamics of z are more sensitive to the bursts.

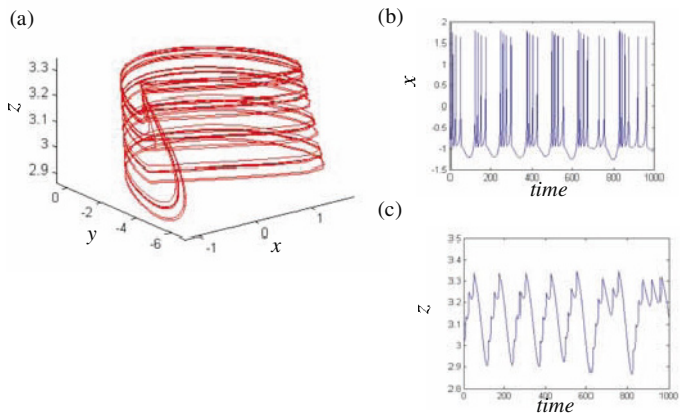


Fig. 23. Phase portrait (a) and time series (b,c) for the Hindmarsh-Rose model. Parameters, I = 3.2; a = 1.0; b = 3.0; c = 1.0; d = 5.0; s = 4.0; x0 = −1.60; r=0.006

Such burst-spiking is of particular interest to classes of thalamic neurons and many cortical neurons during slow-wave sleep, whereby this activity is observed synchronously across the scalp (McCormick & Bal 1997). Interactions between such systems can be introduced by coupling of one of the fast variables, such as

$$\frac{dx_{1,2}}{dt} = y_{1,2} - ax_{1,2}^3 + by_{1,2}^2 - z_{1,2} + I + C\left(x_{2,1} - x_{1,2}\right), \tag{44}$$

where C is the coupling parameter. Two such simulations are shown in Fig. 24, where blue and green time series denote each system. In the top row, with $c = 0.35$, the bursts are coincident but the spikes are often discordant. However with $c = 0.5$, the spikes are also synchronized. This interesting phenomenon, studied in detail by Dhamala et al. (2004) of burst and then spike synchrony, has been observed experimentally.

3.4 Coupled Chaos in a Mesoscopic Model

The Hindmarsh-Rose model introduces an extra term, incorporating a slow calcium current, into a planar model of an excitable neuron. An alternative extension of planar models is to introduce a single variable representing a feedback from an inhibitory neuron Z. The inhibitory and excitatory neurons interact via synaptic currents induced through their mutual connectivity. Such a model takes the form (e.g. Larter et al. 1999),

$$\begin{aligned}\frac{dV}{dt} &= g_{Na} m_\infty(V) \times (V - V_{Na}) + g_K n(V) \times (V - V_K) + g_L \times (V - V_L) \\ &\quad + \alpha_{ne} I + \alpha_{ie} F(Z), \\ \frac{dn}{dt} &= \frac{(n_\infty - n)}{\tau_n}, \qquad \frac{dZ}{dt} = \alpha_{ei} G(V) + \alpha_{ni} I. \end{aligned} \tag{45}$$

Local connectivity is parameterized by the coupling parameters α between inhibitory Z, and excitatory V cells and via input from the external noise

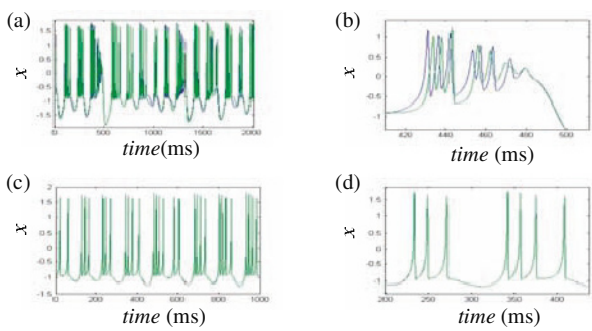


Fig. 24. Coupled Hindmarsh-Rose systems with $c = 0.35$ (top row) and $c = 0.5$ (bottom row)

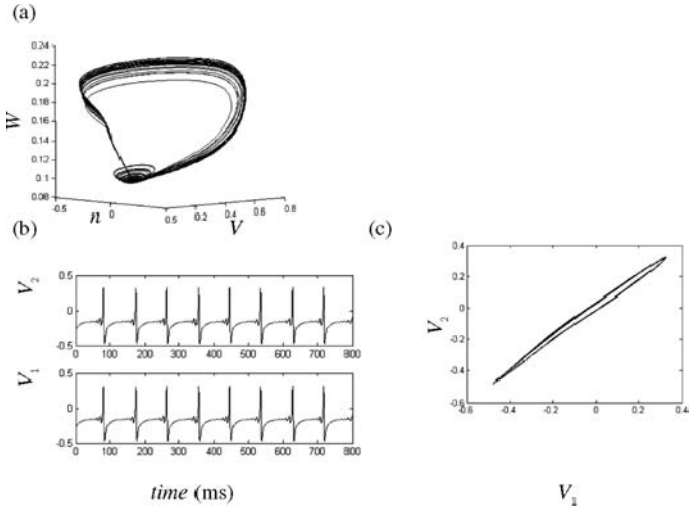


Fig. 25. Generalized chaotic synchronization in a mesoscopic neuronal model. (**a**) Chaotic attractor. The orbits are organized around a manifold that is homoclinic to the unstable spiral (**b**) Time series of excitatory membrane potentials in two coupled systems showing apparent synchronization. (**c**) Their co-evolution shows a smooth manifold slightly off the state of identical synchrony $V_1 = V_2$

term I. The functions F and G model the feedback between the inhibitory and excitatory cells. Within physiologically realistic parameter values, such a system can exhibits chaotic dynamics, as shown in Fig 25 (a), organized around a homoclinic orbit.

Synaptic coupling between the excitatory neurons in two such populations of cells allows construction of a ***mesoscopic*** neuronal model – a system on the intermediate scales between single neurons and the large scale systems considered in the following section. An example of synchronization between two such subsystems is illustrated in Fig. 25 (b-c), where a single parameter in each system has been set with a small mismatch (all other parameters are equal). Whilst the time series appear identical (panel b), a plot of the values of V_1 versus V_2 (panel c) reveals that their co-evolution, whilst close to the diagonal is nonetheless confined to a nearby smooth manifold. This form of non-identical synchronization is known as ***g***eneralized chaotic synchronization (Afraimovich et al. 1986, Rulkov et al. 1995). Further details and examples of more complex behaviors – such as intermittency, scale-free dynamics and travelling waves - can be found in Breakspear et al. (2003, 2005).

This concludes our survey of basic, small-scale neural systems. We hope to have illustrated the power of combining analysis and geometry in elucidating some of the fundamental properties of neurons. We now turn to macroscopic models.

4 From Small to Large Scale Models

Large scale neural network models are thought to be involved in the implementation of cognitive function of the brain (Mesulam 1990; Bressler 1995, 2002, 2003; Bullmore et al. 1996; Mountcastle 1998; McIntosh 2000; Bressler & Kelso 2001; Jirsa 2004; Bressler & Tognoli 2006; Bressler & McIntosh 2007). To understand the neural basis of cognition, theoretical and analytical means must be developed which are specifically targeted to the properties of large scale network dynamics. Such theoretical understanding will also guide the interpretation of the enormous data sets obtained from non-invasive brain imaging. The functional expression of a cognitive operation seems to require the co-activation of certain subnetworks. Such co-activation does not necessarily require a simultaneous activation of all network components, but may be represented in a characteristic spatio-temporal network dynamics with both simultaneous and sequential activations. The properties of the network dynamics will crucially depend on the interconnectivity of the network components and their dynamics (Sporns 2002; Sporns & Tononi 2002, 2007; Jirsa 2004; Beggs et al. 2007). The goal of any large-scale description of neural dynamics is to reconstruct all relevant spatiotemporal dynamics of the neural system while preserving the mechanisms which give rise to the observed dynamics. Large scale models have the implicit assumption to be based upon neurocomputational units, which are more macroscopic than single neurons. This approach is to be juxtaposed with the high-dimensional computation of the full network composed of microscopic complex neurons with dendritic and axonal ion channel dynamics, as well as pre- and postsynaptic processes. Large scale models also bear the promise that they provide insight into the underlying dynamics-generating mechanisms of the network due to their reduced complexity. Finally, large scale models are easier and less time-consuming to be solved computationally. The following sections discuss the various schools of thought in large scale network modeling and characterize these from the perspective of anatomical and functional connectivity, the latter identified with the dynamics of the network.

4.1 Non-reducible Dynamics of Neuronal Ensembles

A large scale model is composed of microscopic units or atoms which do not represent individual neurons, but rather complexes, also referred to as ***neural masses*** (Beurle 1956), capturing the non-reducible dynamics of a set of neurons. Such complexes may either be localized in physical space and defined in a volume element at a location x, or distributed over physical space and are defined functionally (e.g. in K-sets as discussed below (Freeman 1975, 1992)). Though the former is more common, in practice the two variants often coincide due to stronger local connectivity and the resulting co-activations ("what wires together, fires together"). Unlike many subcortical structures, in which neurons are packed into nuclei, the cortical

sheet appears at first sight as a dense homogeneous medium with no obvious demarcation of its components. Corticocortical columns typically consist of 5,000 to 10,000 neurons, macrocolumns contain 10^5 to 10^6 neurons (Nunez 1995). Yet, there are a number of anatomical tracing studies which indicate mutual anatomical and functional parcellation (Szentagothai 1975; Mountcastle 1978). For instance, macrocolumns form functional units in sensory areas with homogeneous tuning properties inside the unit, but sharp differences amongst neighboring units (Mountcastle 1978). For our purposes, the neural mass is a highly connected set of neurons, sharing common input and output pathways and specialized low-level function. The activity of a neural mass (also known as ***neural mass action***) in a large scale model is described by an m-dimensional vector variable $\Psi(x,t) = (\Psi_1(x,t), \Psi_2(x,t), \cdots, \Psi_m(x,t))$ at a discrete location x in physical space and a point t in time. The variable Ψ(x,t) is also referred to as a ***neural population, neural assembly*** or ***neural ensemble activity***. If the distance between neighboring neural masses is infinitesimally small, then the physical space x is continuous and Ψ(x,t) is referred to as a ***neural field***. Since the neural mass action is physically generated by the N neurons within the neural mass, there will be a mapping $\Phi : Z(x,t) \rightarrow \Psi(x,t)$, which unambiguously relates the high-dimensional neuron activity $Z(x,t) = (Z_1(x,t), Z_2(x,t), \cdots, Z_N(x,t))$ to the neural mass action Ψ(x,t). $Z_i(t)$ is the n-dimensional state vector of the i-th neuron with i= 1,...,N. For concreteness, a neural mass may contain N=10,000 neurons with n=2 in case of a FitzHugh-Nagumo neuron model. The situation is shown in the cartoon on the bottom of Fig. 26. Here a cortical sheet is shown which is decomposed into color-coded patches representing neural masses. Within a neural mass the local connectivity of a single neuron is illustrated through the density of its connections (red squares) which decreases with increasing distance. The partial overlap of the neural masses indicates that synaptic connections of a neuron may belong to different neural masses. The critical step in the development of a large scale model occurs through the mapping $\Phi : Z(x,t) \rightarrow \Psi(x,t)$ when the activity $Z(x,t) = (Z_1(x,t), Z_2(x,t), \cdots, Z_N(x,t))$ of a given neural mass is replaced by its neural mass action $\Psi(x,t) = (\Psi_1(x,t), \Psi_2(x,t), \cdots, \Psi_m(x,t))$ where $m << N$. The nature of this relation between neuron activity $Z(x,t)$ and neural mass action Ψ(x,t) will be generally non-trivial and involves a mean-field reduction which will be discussed in the next section. On the top of Fig. 26 the neural network dynamics is now captured by locally coupled neural mass actions Ψ(x,t) assigned to each neural mass at location $x = X_i$. Each neural mass is locally (as indicated at location X_5) and globally (as indicated at location X_1) connected.

Rather than solving the complete network for the state vectors $Z(x,t)$ of all neurons, now the large scale network can be solved using the neural mass action Ψ(x,t) as indicated in the following: A large scale model representation is successful if the large scale model simulation provides the same neural mass

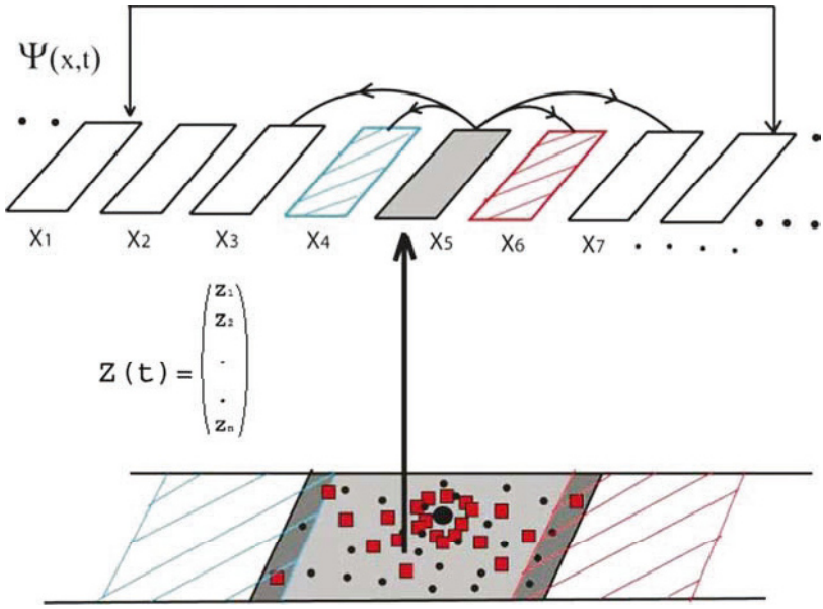


Fig. 26. Coupled neural masses at locations X_i (upper figure) are coupled via local and global pathways. The large scale network dynamics arises from the interactions of the neural mass actions $\psi(X_i, t)$ at locations X_i. The computation of the complete network dynamics based upon the neural state vector Z(t) (lower figure) and neural connectivity (red squares) should ideally yield the same network dynamics as computed from ψ(x,t)

action Ψ(x,t+T) at a future time point t+T as the simulation based upon the complete network dynamics using the microscopic neuronal activity Z(x,t).

$$\Phi : Z(x,t) \rightarrow \Psi(x,t) \xrightarrow{\text{large scale network dynamics}} \Psi(x,t+T) \leftarrow Z(x,t+T)$$

complete network dynamics

In the latter approach, once Z(x,t+T) is computed, it has to be mapped upon to the neural mass action, $\Phi : Z(x,t+T) \rightarrow \Psi(x,t+T)$ to allow for a comparison between the two approaches. The inverse mapping $\Phi^{-1} : \Psi(x,t+T) \rightarrow Z(x,t+T)$ generally does not exist.

4.2 Mean Field Reduction of Neuronal Activity

The ***mean field*** approximation is well-known from statistical physics (see for instance Gardiner 2004). Though its basic assumptions are mostly not rigorously justified, it often provides an astonishingly good qualitative insight into the description of many models. Hence the use of mean field approaches has

a long history in the field of neural networks. The mean field $u(t)$ is generally defined as the statistical expectation value E of a particular state variable. Two mean field approaches exist based on two opposing views of neuronal coding, but of course with many interim shades. The first view holds that the firing rate of a neural mass is relevant for neural information processing. The dissenting view posits that the information is encoded in the interactions among spikes and hence spike correlations must not be ignored (for detailed discussions of neuronal encoding see Koch 1999). In large scale models, neural mass action is mostly expressed by mean fields of firing rate, though also considerable evidence exists that single cells may fire spikes at predictable intervals as long as 200msec with a precision of 1msec (Abeles et al. 1993). The latter is the key observation leading to the theory of synfire chains for cortical processing (Abeles 1991). As of today, it is not clear to what degree the neural system uses firing rate or spike coding mechanisms. Experimental evidence exists for both and accumulates with every day (Koch 1999). In the following we elaborate on the import of both neural coding mechanisms to the field of large scale modeling.

Generally speaking, if the coupling is high enough and the parameter dispersion is sufficiently small, the neurons in the neural mass evolve in time close to each other within phase space (and hence to the mean field), or in other words are synchronized. Note however, that there are exceptions in the network dynamics literature known as oscillator death (see also Campbell 2007), in which the neural mass action becomes zero due to too strong coupling. A synchronized neural dynamics will play a lesser role for extended periods of time during which a large scale synchronization is more likely to indicate pathological network activity such as epilepsy (see Milton et al 2007, Ferree and Nunez 2007). However, the understanding of the conditions leading to the emergence of synchronization will likely be important to understanding the neurocognitive processes such as feature binding (Gray and Singer 1989; Crick and Koch 1990) and multisensory integration (Von Stein et al 1999; Treisman 1996). In fact, the onset of coherent oscillatory activity has been interpreted to be fundamental for the formation of higher-order percepts (Freeman and Skarda 1985; Bressler 1990). In the opposite case for small coupling and greater noise strength, the elements of the population move incoherently and eventually their positions average out. Here the asymptotic dynamics of the mean field is mostly characterized by the fluctuations and the mean firing rate. Between these two limit cases, complex behavior arises and can be addressed starting from either end of the limit.

Fluctuation dominated network dynamics and firing rate models

For small enough and sparse couplings, as well as sufficient noise within the neural mass, the neuronal action potential generations and the connectivity within the neural mass can be assumed to be independent. Under these conditions, all spike correlations will be destroyed and a ***firing rate model***

becomes a valid representation of neural mass action (Abbott & van Vreeswijk 1993; see Cessac & Samuelides 2006 for a review). In the limit of large neuron numbers within the mass, N→∞, and low firing rates, the total spike train, obtained by summing over the spike trains from all neurons within the mass, will be a Poisson point process with a common instantaneous firing rate ρ(x,t). Equivalently, the synaptic input I_s to a single neuron can be approximated by an average firing rate ρ(x,t) plus a fluctuating Gaussian contribution. As a consequence the joint probability distribution factorizes and a complete description of the neural mass action is obtained in terms of the first and second order statistical moments. Two further more subtle distinctions can be made. Either the firing rate ρ(x,t) plus Gaussian noise is used as synaptic input and the neural mass action is described by the average value of neural activity (the mean field) $u(t) = E[Z_i(t)]$ and the variance $v(t) = E\left[Z_i(t)^2\right] - u^2(t)$. Such finally results in Fokker-Planck approaches which describe the time evolution of the probability *P(Z(x,t),t)* to find a neuron at x and t in the state Z (Amit & Brunel 1997; Brunel 2000; Brunel & Hakim 1999; Cai et al. 2006). Alternatively, the neural mass action can be expressed directly by the mean firing rate $u(t) = E[\rho(t)]$ and its variance $v(t) = E\left[\rho(t)^2\right] - u^2(t)$ (Abbott & van Vreeswijk 1993; Nykamp & Tranchina 2000, 2001; Eggert & van Hemmen 2001). Note that we dropped the explicit dependence on x to simplify our notation. The mean field variables u and v define the 2-dimensional population vector $\Psi(x,t) = (u(t), v(t))$ at the location x. As long as the independence condition within the neural mass holds, the reduced dynamic description through Ψ(x,t)=(u(t),v(t)) is exact. The mean firing rate shows a sigmoid behavior as a function of the synaptic input, which can be intuitively understood as follows: a neural mass shall consist of independent neurons of which each displays a sharp onset of firing at a threshold value Θ (see Fig. 27).

The thresholds are independent and hence have a Gaussian distribution. The mean firing rate of the neural mass then becomes the well-known sigmoid function and has been carefully parameterized from experimental data of the olfactory bulb (Freeman 1975). Is the independence condition violated though

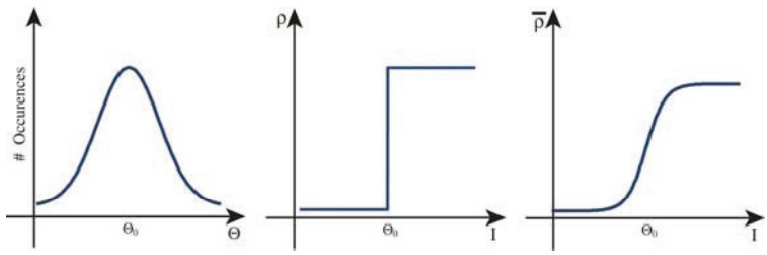


Fig. 27. Left: Gaussian distribution of activation thresholds within a neural mass. Middle: Sharp activation function (firing rate) of a single neuron acting as a threshold element. Right: Mean firing rate obtained from averaging all firing rates within a neural mass

and correlations are introduced, for instance through correlations within the connectivity weights via learning, the mean field approximation breaks down. Related in spirit to Fokker-Planck approaches, Ventriglia proposed a phenomenological kinetic theory for the study of the statistical properties of neural mass action (Ventriglia 1974, 1978). The kinetic equations capture the time course of the distribution function of the total excitation of a neural mass. The neurons in the mass are characterized by a level of inner excitation which changes when impulses are emitted. The impulses move freely within the neural mass and may be absorbed by other neurons changing their inner excitation level (see also by Gröbler et al. 1998; Barna et al. 1988 for extensions of the kinetic approach).

Synchronized network dynamics in population models

For strong coupling strengths and low level of noise within the neural mass, a different but complementary approach holds using the perfectly correlated state $\mathrm{Z}(t)$, that is $Z_1(t) = Z_2(t) = \cdots = Z(t)$. Or in other words, a special case of spike timing is considered: all neurons are synchronized and show the same dynamics $Z(t)$. DeMonte, d'Ovidio & Mosekilde (2003) proposed a method by which the mean field dynamics of a neural mass can be described by a low-dimensional population vector under conditions of global coupling and coherent neural mass action. Global coupling means that each neuron in the neural mass feels the same mean field activity. Their method applies to neural masses of any size and any type of intrinsic dynamics, as well as parameter dispersion. For example, if the underlying neuron model for $Z_i(t)$ is a FitzHugh-Nagumo Model, then the population vector $\Psi(x,t) = (\Psi_1(x,t), \Psi_2(x,t), \Psi_3(x,t), \Psi_4(x,t))$ is 4-dimensional where $\Psi_1(x,t), \Psi_2(x,t)$ describes the activity of an average FitzHugh-Nagumo neuron and $\Psi_3(x,t), \Psi_4(x,t)$ measures the dispersion of both parameter and phase space. If the neurons desynchronize too much, then the approach of DeMonte et al. (2003) will fail by definition. If, after loss of synchrony, multiple clusters of coherent activity emerge in phase space instead, then it is possible to describe the neural mass action through multiple mean fields. Each of these mean fields captures a single cluster dynamics (Assisi, Jira & Kelso 2005). In cases of parameter dispersion, such emergence of cluster dynamics is common and well-suited for the approach by Assisi et al. (2005). If the constraint of global connectivity within the neural mass is dropped, richer dynamic phenomena become possible such as the appearance of spiral waves (Chu et al. 1994; see Milton 1996) and will be discussed in the next sections. Freeman (1975,1987) proposed another classification of neural mass action which allows spike correlations to be considered. He originally classified the activity of the neural masses into classes named K0, KI and KII sets (K for Katchalsky) according to their functional architecture. K-sets are composed of elements which affect the nature of dynamics including physical components such as the interconnected neurons, the neurochemical environment, etc., but also purely

functional components such as the connection topology, the input structure, etc. K0 sets represent the simplest functional architecture which can be viewed as the ensemble average of the activity of independent but similar neurons. In their simplest forms, KI sets are equivalent to two coupled K0 sets, KII sets are composed of K0 and KI sets. However, they are more generally defined and are in principle not always reducible to lower order K sets. In this notation, a K0 set corresponds to the 1-dimensional and hence scalar activity of a neural mass, Ψ(x,t), whereas KI and KII sets correspond to higher-dimensional vectors Ψ(x,t).

4.3 Composition of Neural Masses to Large Scale Models

Neural mass models sacrifice realism for a more parsimonious description of the key mechanisms of large scale dynamics. The benefit lies in the possibility of emulating non-invasively obtained brain imaging data such as EEG and MEG. Neural mass models (Beurle 1956; Lopes da Silva et al. 1974; Freeman 1975; Nunez 1974, 1995; van Rotterdam et al. 1982; Jirsa & Haken 1996, 1997; Jirsa et al. 1998, 2002; Robinson et al. 1997, 2002, 2001; Tagamets & Horwitz 1998; Steyn-Ross et al. 1999; Valdes et al. 1999; David & Frison 2003; Breakspear et al. 2006) are based upon this approach. Much of the complexity of the signals arises from the coordination of the interconnected neural masses rather than the intrinsic dynamics of the microscopic unit, the neural mass, of the large scale network. A neural mass at location x is locally connected to its neighboring neural masses and globally connected to far distant neural masses at locations x'. In the following, physical space is always assumed to be one-dimensional, $x \in \Re$, but the mathematical treatment formally extends trivially to two and three dimensions. Note that though the formal extension to higher dimensions is not difficult, new dynamic network phenomena such as spirals may emerge due to the higher dimension (see Nunez (1995) for a discussion of spherical geometries). If the network dynamics described by (17) were linear, then the mapping $\Phi : Z(x,t) \rightarrow \Psi(x,t)$ would result in the following large scale dynamics for the neural mass action Ψ(x,t) with $Q = N$ and $S = H$

$$\frac{d\Psi(x,t)}{dt} = Q(\Psi(x,t)) + \int_{\Omega} \int_{-\infty}^{t} h(x-x')\, S(\Psi(x-x', t-t'))\, dt' dx'. \quad (46)$$

However, in general the intrinsic dynamics N and the activation function H are nonlinear and residual terms arise which are here notationally absorbed in Q and S. The intrinsic, sometimes also called endogenous, dynamics N of the neural mass action is defined by the temporal evolution of Ψ(x,t) in absence of all incoming signals including the connections to other neural masses. In the following we will discuss representative models from this line of approach and characterize the various entry points towards large scale network modeling.

We place particular emphasis on the functional effects that the variation of structural properties, such as local and global connectivity and time delays, implies.

Amari's Neural Field Model 1977

A classic paper on networks with no delay and symmetric and translationally invariant connection topologies is Amari's study of neural fields (Amari 1977). Amari discussed spatially and temporally continuous fields Ψ(x,t) with local fixed point dynamics as intrinsic dynamics. Then the field equations may be written as

$$\tau \frac{d\Psi(x,t)}{dt} = -\Psi(x,t) + \int_{\Omega} h(x-x')\, S(\Psi(x',t))\, dx' + c + s(x,t)\,. \qquad (47)$$

where S is strongly nonlinear, typically the Heaviside function, and $h(x-x')$ is excitatory for proximate connections and inhibitory for greater distances (see Fig. 28 and 29). $s(x,t)$ denotes external input and c a constant resting potential and background activity.

In this type of scalar neural fields, oscillations are not possible, but locally excited regimes may exist and self-sustain with no input $s(x,t) = 0$, which is believed to be a candidate for the neuronal basis of working memory (Amit 1989). If input is provided, then the locally excited regions travel in the direction of increasing field value Ψ(x,t) until they get pinned at the stimulus location.

If several stimuli are provided, then the details of stimulus location and the presence of already excited local regions will determine the typically

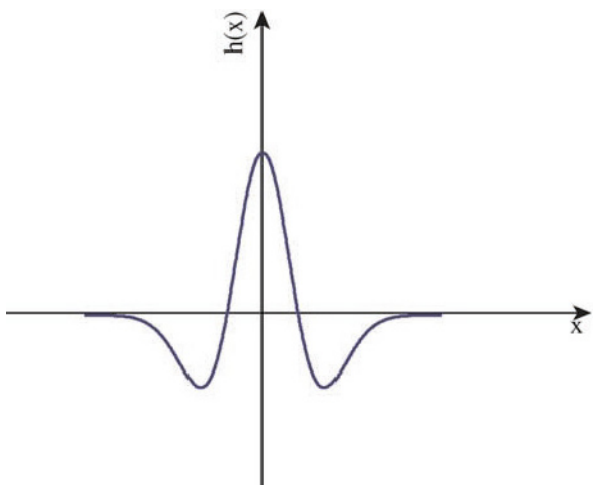


Fig. 28. Distribution function is plotted which captures local excitatory and lateral inhibitory connectivity

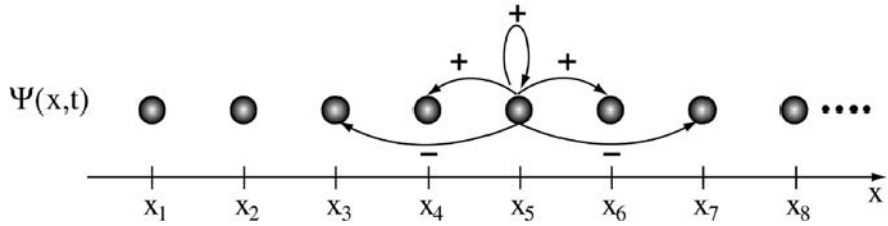


Fig. 29. The characteristic connectivity of an Amari field reflects local excitation and lateral inhibition

multi-stable final network dynamics. Characteristic examples are shown in Fig. 30(a), b and c.

In all cases, the final stationary network state will be a fixed point attractor. It was these properties, which attracted the attention of neural modelers who applied these fields to a variety of phenomena ranging from working memory (Amit 1989) to motor movement preparation (Erlhagen & Schöner 2002). If two or more layers are coupled (Amari 1977), then a more complex dynamics arises allowing for oscillatory and traveling wave phenomena.

The Neural Field Models of Wilson & Cowan (1972, 1973) and Nunez (1974)

Hugh Wilson & Jack Cowan (1972, 1973) and Paul Nunez (1974) independently considered twocomplementary approaches, of which each is based upon

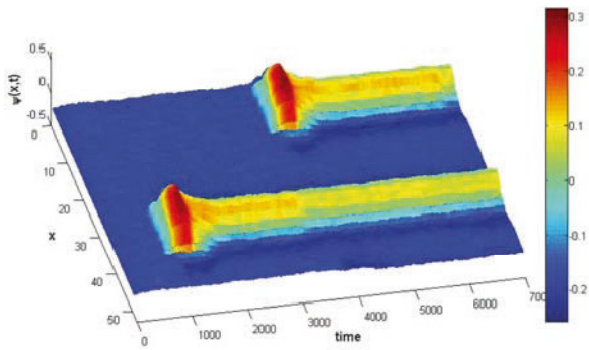


Fig. 30(a). The space-time diagram of an Amari field is shown. Initially the neural field is not excited, then a stimulus is introduced around 1000ms at location x = 35 (space is in arbitrary units). At stimulus offset around 1300ms, the neural field sustains its local excitation. At a later time point, another stimulus is introduced at x = 10 for 300ms. Here the neural field also persists after stimulus offset. Such persistent activity serves a s a simple model for working memory

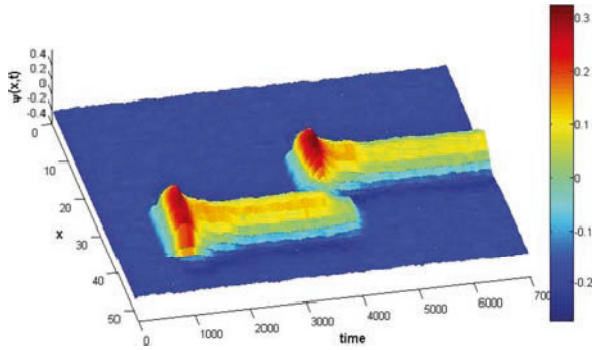


Fig. 30(b). The same situation is shown as in Figure 30a, only the second stimulus is provided closer in space, x = 25, to the first stimulus and annihilates the excitation at x = 35. The second local excitation persists unaltered

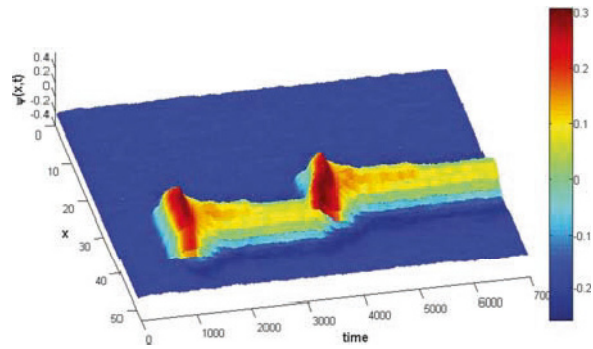


Fig. 30(c). The same situation is shown as in figure 30b, only the second stimulus is now provided even closer in space, x = 30, to the first stimulus than before. This time it does not annihilate the excitation at x = 35, on the contrary, both excitations move towards each other and merge into one excitation. In the figure, it appears that the excitation at x = 35 moves more than the other, which is true

two sets of locally coupled neural masses of inhibitory and excitatory neurons. Wilson & Cowan considered the firing rate as the neural mass action; Nunez considered synaptic action which is the proportion of active synapses at time t and linearly related to dendritic currents. The firing rate of neural masses has been referred to as pulses and the synaptic action as waves (Freeman 1975). Jirsa & Haken (1996, 1997) showed that both models are equivalent and can be transformed into each other using so-called pulse-wave and wave-pulse conversions, which are independently experimentally accessible (Freeman 1975). Both models consider time delays via propagation. Delays are absent in Amari's model and hence constrains the latter's applicability in a biologically realistic scenario to small patches of cortical tissue. Time delays are of increasing importance, the larger the scale of the network is.

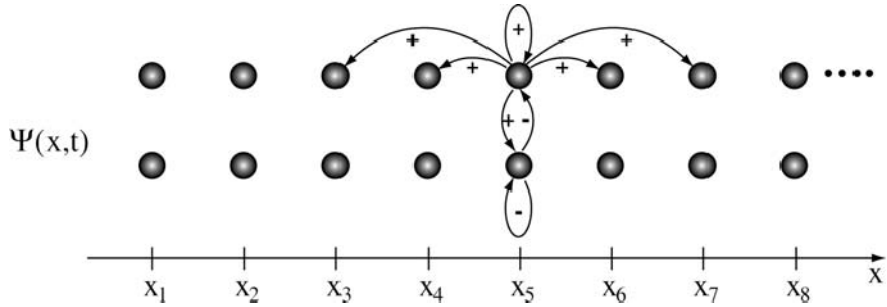


Fig. 31. The characteristic connectivity of Nunez's population approach reflects local and global excitation, as well as local inhibition

In Nunez's early work (1974), his focus was on identifying the dispersion relations of the linearized neural field dynamics given specific distributions of intracortical and corticocortical fiber systems. The intracortical fiber system is constrained to the gray matter and its axons make connections within a few millimeters; the corticocortical fiber system constitutes the white matter and connects areas across the entire cortex with axonal lengths of several centimeters in the human (Abeles 1991; Braitenberg & Schüz 1991), in some cases reaching lengths of up to 15 to 20 centimeters (Nunez 1995). The excitatory synaptic action Ψ_1(x,t) and inhibitory synaptic action Ψ_2(x,t) compose the neural mass action and define one excitatory and one inhibitory layer (see Fig. 31). The dynamics of the two-dimensional neural field is governed by the following equation

$$\frac{d\Psi\,(x,t)}{dt} = -\Psi\,(x,t) + s(x,t) + \int_{\Omega}\int_{0}^{\infty} h\,(x-x',v)\,S\left(\Psi\left(x,t-\frac{|x-x'|}{v}\right)\right)dvdx', \tag{48}$$

where Ψ(x,t) = (Ψ_1(x,t), Ψ_2(x,t)), s(x,t) is the input to the two layers, h(x−x',v) defines a matrix describing the distribution of axonal fibers, S is the sigmoid firing rate and Ω defines the spatial extent of the neural sheet. Due to the finite transmission speed v, there is a time delay $\frac{|x-x'|}{v}$ via propagation.

The connectivity function h(x−x',v) is a 2 by 2 matrix, since Ψ(x,t) is a 2-dimensional vector field, and considers both intracortical and corticocortical fibers collapsed into one distribution function. The synaptic influence is assumed to diminish in proportion to its density, in particular Nunez extrapolated h from mouse data (Nunez 1995) to assume an exponential form,

$$h(x) = \exp\left(-|x|/\sigma\right)/2\sigma, \tag{49}$$

as illustrated in Fig. 32, with the rate of drop-off captured by the parameter σ.

The inhibitory connectivity is of short range and the excitatory connectivity is of long-range since the latter is dominated by the corticocortical fiber

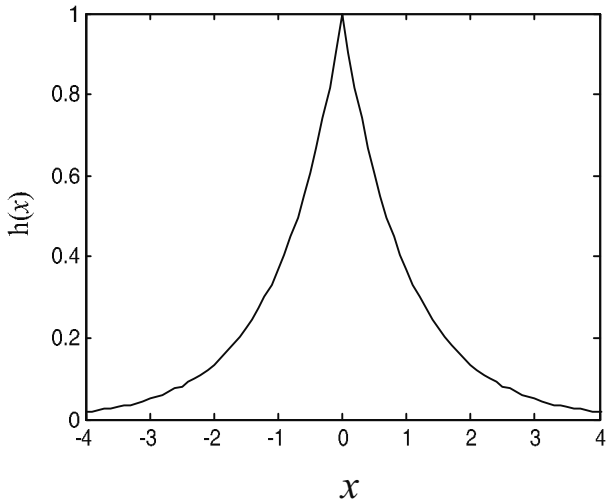


Fig. 32. Exponential coupling function

system. It is notable that Nunez's approach represents the first attempt to use structural information to constrain the neural field dynamics for the large scales observed in encephalographic measurements. His consideration of the corticocortical fiber system within the integral kernel of (49) has influenced much of the later research in the field of EEG and MEG (Jirsa & Haken 1996,1997; Wright and Liley 1996; Jirsa et al. 1998, 2002; Robinson et al. 1997, 2001; Steyn-Ross et al. 1999 ; Breakspear et al. 2006) and lead to the development of neural field dynamics for large scale systems.

Wilson & Cowan (1972) initially considered the interaction of two populations of excitatory and inhibitory nature characterized by their firing rates Ψ_1(t) and Ψ_2(t). An der Heiden (1980) showed nicely the connection between the local Wilson-Cowan population model (1972) and the McCulloch-Pitts model (1943). Later Wilson and Cowan (1973) extended their model to two layers of coupled neural fields Ψ_1(x,t) and Ψ_2(x,t) (see Fig. 33) obeying the following equation

$$\frac{d\Psi\left(x,t\right)}{dt} = -\Psi\left(x,t\right) + S\left(\int_{\Omega} h\left(x-x'\right)\Psi\left(x,t-\frac{\left|x-x'\right|}{v}\right)dx' + s(x,t)\right), \tag{50}$$

where we use the same notation as in the Nunez model. Various dynamic phenomena were found as a function of the connectivity h including steady network states, standing and traveling waves. An emphasis was placed on the spatial localization of activations, which functionally necessitated the constraint that inhibitory connections are of longer range than excitatory interactions. Such is in analogy to Amari's model and is anatomically reflected in the longer axons of inhibitory interneurons. This constraint, however, requires

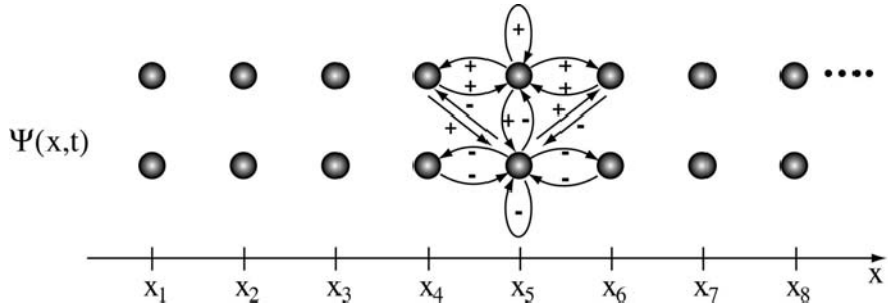


Fig. 33. The characteristic connectivity of Wilson & Cowan's coupled population approach reflects local excitation and inhibition with various degrees of laterality

the axons not to leave the gray matter and clearly limits the application of these neural fields to local area networks.

The Neural Field Model of Jirsa & Haken (1996) and Wave Equations

Based on first principles using pulse-wave and wave-pulse conversions, Jirsa & Haken (1996, 1997) developed a neural field approach (see for early accounts of neural field theories Griffith 1963, 1965) targeted specifically towards large scale phenomena as observed in EEG, MEG. Initially based on two locally coupled neural masses of excitatory and inhibitory neurons, the action of the inhibitory neural mass is absorbed into an effective excitatory neural mass action $\Psi(x,t)$. This reduction is possible under the assumption that the intrisic dynamics of the neural mass is negligible and relaxes instantly to its steady state, i.e. the neural mass action displays a fixed point dynamics. Then the network dynamics will exclusively be determined by the connectivity and its time delays and captured by an equation equivalent to (48), but with a scalar connectivity function h(x−x') (see Fig. 34). As in the Nunez model, the connectivity includes local intracortical connections and global corticocortical projections. As a first approximation, h(x−x') is assumed to be translationally invariant and follows an exponential decay as plotted in Fig. 32. Under these

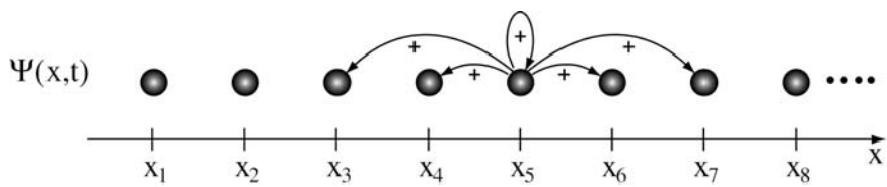


Fig. 34. The characteristic connectivity of the Jirsa-Haken wave equation emphasizes excitatory long range connectivity after elimination of the local inhibitory effects. The latter are captured in an effective neural mass action $\Psi(x,t)$

conditions, Jirsa & Haken showed that the integro-differential (48) is equivalent to the following partial differential equation in one physical dimension:

$$\frac{\partial^2 \Psi(x,t)}{\partial^2 t} + 2\omega_0 \frac{\partial \Psi(x,t)}{\partial t} - v^2 \frac{\partial^2 \Psi(x,t)}{\partial^2 x} + \omega_0^2 \psi(x,t) = \omega_0(\omega_0 + \frac{\partial}{\partial t}) S(\Psi(x,t)) \tag{51}$$

where $\omega_0 = v/\sigma$, v is the transmission speed along myelinated axons and σ the mean fiber length. Early accounts of wave phenomena in EEG and their discussion in the context of wave equations can be found in (Nunez 1995).

The Jirsa-Haken wave equation (51) approximates various connectivity functions of large scale networks in the limit for long waves, or, in other words, large scale activity patterns. If the slope of the sigmoid function S increases beyond a threshold, then the rest state becomes unstable and undamped wave propagation occurs. Below the threshold damped wave propagation exists. Steven Coombes and colleagues (2003) discuss the effects of connectivity strengths which do not decrease with increasing distance, but rather remain constant within a finite regime. In this case, it is not sufficient to describe the spatiotemporal dynamics by a local partial differential equation as in (51), but non-local delayed terms arise (see Coombes 2005 for a review). Wright and colleagues introduced much physiological detail and were able to address issues of rhythm generation (Wright & Liley 1996), as well as clinical aspects such as hysteresis phenomena in anesthesia (Steyn-Ross et al. 1999). Robinson and colleagues introduced expressions for the corticothalamic loop into the Jirsa-Haken equation (see next section) and included dendritic dynamics while implementing detailed physiologically realistic parameter ranges (Robinson 1997, 2001). Frank and colleagues developed a Fokker-Planck approach to the Jirsa-Haken equation which captures the time evolution of the stochastic properties of the neural fields (Frank et al. 1999, 2000). Applications to encephalograpic data can be found in (Jirsa and Haken 1997; Jirsa et al. 1998, 2002; Fuchs et al. 2000; Liley et al. 2002; Jirsa 2004b; Robinson et al. 2004, 2005; Breakspear et al. 2006).

The Inclusion of the Thalamocortical Loop into Neural Fields (Robinson 2001)

In 1997 Robinson et al. presented an equivalent derivation of the Jirsa-Haken equation considering effects of dendritic dynamics and added the important extension of the thalamocortical loop in 2001 (see Fig. 35). The inclusion of the thalamocortical interactions proved to be crucial to reproduce the essential spectral properties observed in scalp topographies. Robinson and colleagues preserve the neural field as a vector field $\Psi(\mathrm{x,t}) = (\Psi_1(\mathrm{x,t}),\ \Psi_2(\mathrm{x,t}))$ of excitatory and inhibitory neural masses and write the following equations

$$\begin{aligned} \frac{\partial^2 \Psi(\mathbf{x},t)}{\partial^2 t} + 2\omega_0 \frac{\partial \Psi(\mathbf{x},t)}{\partial t} - v^2 \frac{\partial^2 \Psi(\mathbf{x},t)}{\partial^2 x} + {\omega_0}^2 \psi(\mathbf{x},t) \\ = {\omega_0}^2 \rho(\Psi(\mathbf{x},t), \Psi_{th}(t - \tau/2)) \end{aligned} \tag{52}$$

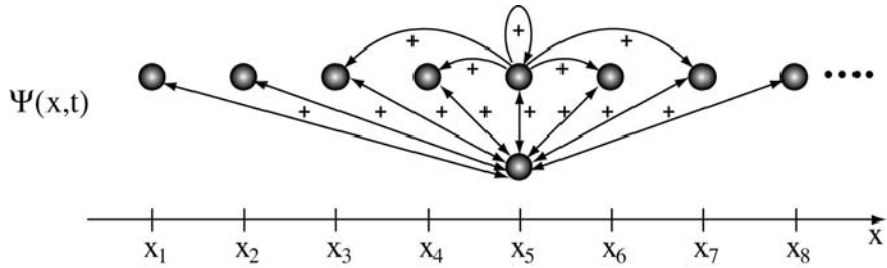


Fig. 35. Robinson et al. (2001) capture corticothalamic effects contributing to neural field dynamics. The effect of excitatory and inhibitory influences is collapsed into the upper row

for the dynamics of the neural field. The critical step is that the sigmoid function $\rho(\Psi(\mathbf{x}, t))$ does not only depend on the neural fields Ψ_1(x,t), Ψ_2(x,t), but also receives time–delayed thalamic input $\Psi_{\mathrm{th}}(\mathrm{t} - \tau/2)$.

The thalamic action Ψ_{th}(t) is governed by the following differential equation

$$\frac{\partial^2 \Psi_{th}(t)}{\partial^2 t} + (a+b)\frac{\partial \Psi_{th}(t)}{\partial t} + ab\Psi_{th}(t) = \mathrm{input}(\Psi(x, t - \tau/2)) \qquad (53)$$

where the cortical input to the thalamus also undergoes a delay $\tau/2$ via propagation resulting in a effective delay τ of the total corticothalamic loop. Computer simulations of equations (52) and (53) provide representative EEG power spectra as shown in Fig. 36.

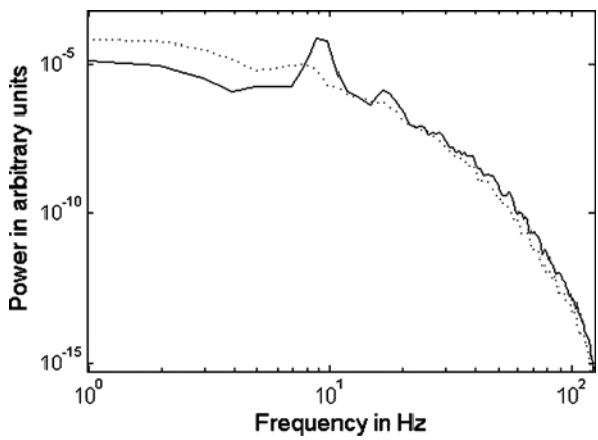


Fig. 36. Power spectra from the Robinson model of corticothalamic activity in eyes closed (solid) and eyes open (dashed) resting states. The increase of low frequencies in the eyes open condition reflects increased corticocortical gain, whereas the increased alpha (10 Hz) peak in the eyes closed condition reflects increased corticothalamic gain

Extensions and Limitations of Large Scale Models

Large scale systems are characterized by an anatomical connectivity with massively parallel and serial, hierarchical structures, as well as time delays due to signal transmission. Such architecture produces an interareal connection topology, which is patchy as observed by Braitenberg & Schüz (2001) and results in a heterogeneous connectivity. Yet it has been approximated in various attempts by a homogeneous connectivity with a larger extension (see Fig. 37). The approach uses a larger mean path length and hence effectively mixes functionally the intracortical and corticocortical fiber systems. Research of this kind has successfully reproduced various large scale characteristics of activity including the dispersive properties of the cortex (Nunez 1995) or global EEG power spectra (Robinson 2001); it also shows promise in situations of highly symmetric functional connectivity (Jirsa et al. 1997, 1998; Fuchs et al. 2000). However, to this date, it has not been shown rigorously under what conditions the homogeneous approximation holds.

Mallot and colleagues (1989, 1996) discussed in a series of papers a conceptual framework in which, rather than just mean fields, local networks communicate across distances. These local networks have an intrinsic fixed point dynamics, but exchange information via time-delayed pathways. Mallot and colleagues applied this approach to examples of the thalamocortical loop (Mallot et al. 1996) and for the geniculate-striate pathway of the visual system (Mallot et al. 1989). Similarly, discretely coupled local networks incorporate time delays in the connecting pathways and absorb all local dynamics within a set of coupled neural masses (Freeman 1975, 1992; David & Frison 2003). Jirsa & Kelso (2000) studied the neural field dynamics of the Jirsa-Haken equation in which a heterogeneous pathway is included (Fig. 38). Such a two-point pathway connects the neural masses at locations X_2 and X_8 which are embedded into a continuous sheet with local connections only. This

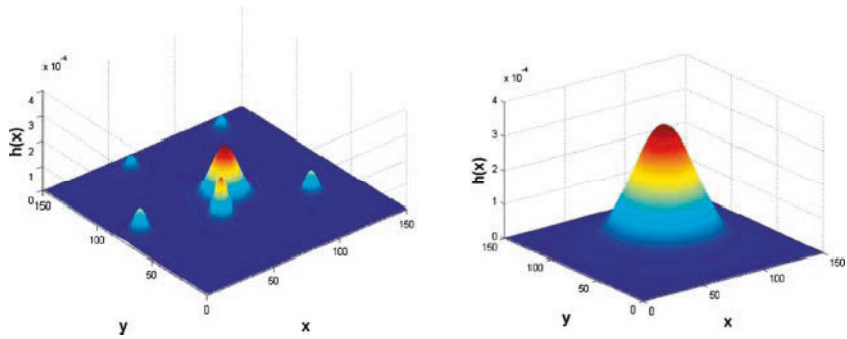


Fig. 37. Left: Neural connectivity has local intracortical symmetric components (homogeneous) and patchy corticocortical components (heterogeneous). Right: Approximation of the real local and global connectivity by a symmetric connectivity function with an average path length

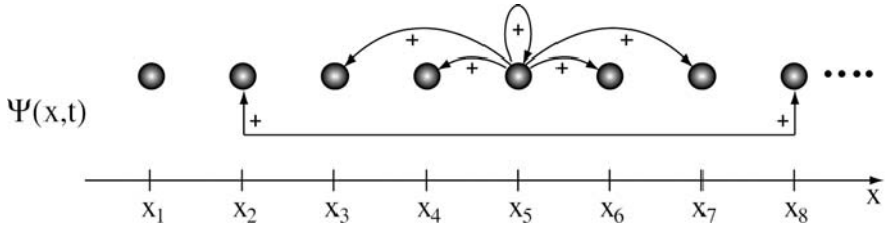


Fig. 38. Realistic connectivities are characterized by translationally invariant local connections and translationally variant global connections. The basic model for the study of the interplay of local and global interactions is the embedded two-point connection in a locally connected neural network as shown in the figure

connectivity identifies the basic toy model for the study of local and global interactions. The change in the connection topology destabilizes the initial stationary dynamics and the system undergoes a transition to a new stationary state via a Hopf bifurcation. Detailed bifurcation diagrams are given in (Jirsa & Kelso 2000) in which the spatiotemporal reorganization is characterized as a function of the length of the two point connection.

Minor changes in the location of the terminals or the system parameters, such as the homogeneous or heterogeneous transmission speeds, may result in qualitatively different global neural field dynamics. As an example, in Fig. 39 a stimulus is introduced in the neighborhood of a terminal of a two-point

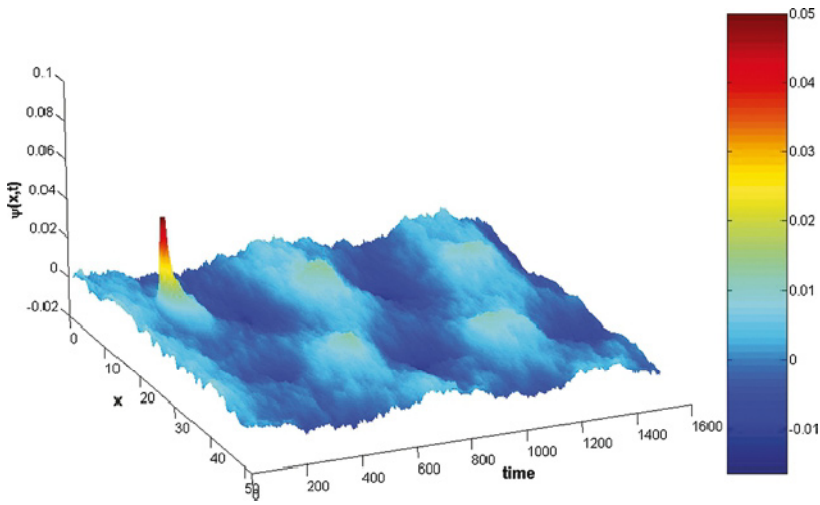


Fig. 39. A neural field following Jirsa and Haken (1996) with an embedded two-point connection at x=10 and x=40 is established. In the neighborhood of x=10, a brief stimulus excites the neural sheet locally and the neural field reorganizes globally in a large scale transient wave which damps out after a sufficiently long time (not shown here)

connection (based at x=10 and x=40), then the excitation of the neural field travels through the continuous sheet, but also transmits a signal via the heterogeneous pathway. A transient wave dynamics is observed on the global system scale and damps out after a sufficiently long enough time. With no heterogeneous connection, only a local excitation at x=10 would have been observed. Similarly, with no heterogeneous connection and with two stimuli at terminal sites x=10 and x=40, only two local excitations would have been observed, but no large scale organization as observed in Fig. 39.

5 Conclusion

The neurosciences have historically leaned strongly towards empiricism – a tradition which continues today. However, mathematical formalisms of dynamical phenomena have provided extraordinary explanatory and unifying insights in the physical sciences. The emerging advances in computational neurosciences, particularly with respect to brain connectivity, suggest that they will also come to play an important role in the brain sciences. The cross-fertilization of dynamical systems theory (see also the Chapters by Campbell, Horowitz & Husain and Stephan & Friston), graph theory (Sporns & Tononi), basic physics (Ferree & Nunez), and methodological advances in neuroimaging (Darvas & Leahy, Fuchs) will hopefully underpin advances which do not merely reduce problems in neuroscience to problems already solved in other fields, but instead allow those properties of the brain that are unique to inform novel and specific discoveries. We see this blending of universality and specificity as absolutely crucial. Too much of the former will yield simplifications that lose what is required of a system in order to look (and function) like a brain. Conversely, too much specific detail yields volumes of descriptive data that adds little to our understanding of the underlying principles of brain function.

In this chapter, we have overviewed developments in the field of dynamical neural modeling across several scales of magnitude – from the microscopic conductance models of bifurcating neurons, (briefly) through systems of coupled chaotic oscillators at the mesoscopic scale to models of large scale neural networks whose behavior generates the electroencephalographic and neuroimaging data that is acquired non-invasively from human subjects. Evidence of computationally significant processes has been documented in data sets from across this spectrum of scales – i.e. from the single cell to the whole brain. An open and important question then is the relationship between activity at different temporal and spatial scales (Churchland and Sejnowski 1992). A possible answer could be that the macroscopic dynamics is an epiphenomenon – that is, a summed output of dynamics that can only truly be modeled at the neuronal level. However, this approach cannot be reconciled with the successes of large-scale models, which engage the brain at macroscopic scales only, to provide descriptive explanations of neuroscience data. That is, as discussed in

Sect. 4 of the present chapter, a mean field reduction of the present state of the system is able to predict its future states. This suggests that synchronizing processes are able to enslave many of the (small scale) degrees of freedom into dynamical structures at larger scales whose behavior is then – to some degree – determined by the state of the system *at that scale*. Whilst the large scale processes that are sustained by such processes inevitably influence the dynamics of the small scale units, it also remains possible that small scale events – such as critical sensory inputs – are able to rapidly influence the behavior of the system as a whole.

Whilst such considerations preclude a purely reductionist approach, an adequate explanatory framework remains elusive. The situation may be analogous to a heated magnet that it is close to the Curie temperature (above which it loses the ability to be magnetized): The magnetic fields *are purely* an outcome of the dipoles of spinning electrons. Yet the spinning electrons are also strongly influenced by the larger-scale magnetic fields. Below the Curie temperature, the fields are sufficiently strong to overcome stochastic fluctuations of individual spin directions. Above the Curie temperature the emergent fields are insufficient in strength to enslave the electron dipoles and the metal cannot hold a macroscopic field. However, at the Curie temperature, there is just a sufficient degree of coherence at any given scale to overcome the stochastic fluctuations at the next smaller scale. However, fluctuations at a small scale are able to transiently cascade to larger scales, a phenomena known as *criticality* and exhibit scale free fluctuations. Or perhaps even more attractive are the *spin glass* systems where – in addition to these processes – there exist disordered structures embedded in the system which preclude a perfectly ordered system even at low temperatures. Many of such spatiotemporal pattern formation phenomena and their underlying mechanisms have been understood in the framework of *Synergetics*, a field pioneered by Hermann Haken (1983, 1999).

Whilst such arguments have an attractive appeal, we should bear in mind our own warning that the brain is *not* just another complex physical system – such as a heated metal – even one with embedded impurities! There exist additional complexities that are surely important to brain function. One such critical difference is that there do exist structures across spatial and temporal scales prior to the emergence of dynamically driven scale-free (and scale-specific) fluctuations. Is it possible that the interaction between scale-specific processes across a hierarchy of scales is somehow optimal? Fusi *et al.* (2005) have shown how a hierarchy of synaptic processes – each with characteristic time scales – can interact in order to optimize memory retention (upgrading new memories) and storage (maintaining selected memory for long periods of time). Breakspear & Stam (2005) modeled the interaction between scale-free dynamics and multiscale spatial architectures by defining dynamical systems on different wavelet subspaces, and with cross-scale coupling between subspaces. This would potentially allow for a recursive relationship between small and large-scale dynamics.

Such observations hopefully reflect the challenge of fusing the universal with the specific as an emerging frontier in neuroscience research.

Acknowledgements

The authors wish to acknowledge helpful comments by SA Knock, PL Nunez, PA Robinson and E Tognoli and funding by ATIP (CNRS), Brain NRG JSMF22002082, NH&MRC, Australia and ARC Australia.

References

Abeles M (1991) Corticonics. Cambridge University Press

Abeles M, Bergman H, Margalit E, Vaadia E (1993) Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. J. Neurophysiol.70, 1629–1638

Abbott LF, van Vreeswijk C (1993) Asynchronous states in a network of pulse-coupled oscillators. Phys. Rev. E 48, 1482–1490

Abraham RH, Shaw CD (1988) Dynamics – The Geometry of Behavior. Part Four: Bifurcation Behavior. Addison-Wesley: Redwood City.

Afraimovich V, Verichev N, Rabinovich M (1986) Stochastic synchronization of oscillation in dissipative systems. Radiophysics and Quantum Electronics 29, 795–801.

Amari S (1977) Dynamics of pattern formation in lateral-inhibition type neural fields. Biol. Cybern. 27, 77–87

Amit DJ (1989) Modelling Brain Function. New York, Cambridge University Press

Amit DJ, Brunel N (1997) Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. Cerebral Cortex 7, 237–252

An der Heiden U (1980). Analysis of neural networks. In Lecture Notes in Biomathematics (S Levin, ed), volume 35. Springer-Verlag, New York

Arbib M, Érdi P (2000) Structure, Function, and Dynamics: An Integrated Approach to Neural Organization. Behavioral and Brain Sciences 23, 513–571

Ashwin P, Buescu, J, Stewart, I (1996) From attractor to chaotic saddle: A tale of transverse stability. Nonlinearity, 9: 703–737.

Ashwin P, Terry J (2000) On riddling and weak attractors. Physica D, 142, 87–100.

Baker GL, Gollub JP (1990) Chaotic Dynamics: An Introduction. Cambridge University Press

Barana Gy, Gröbler T, Érdi P (1988) Statistical model of the hippocampal CA3 region I. The single-cell module: bursting model of the pyramidal cell. Biol. Cybern. 79, 301–308

Beggs JM, Klukas J, Chen W (2007) Connectivity and dynamics in local cortical networks. This Volume.

Beurle RL (1956) Properties of a mass of cells capable of regenerating pulses. Philos. Trans. Soc. London Ser. A240, 55–94

Braitenberg V, Schüz A (1991) Anatomy of the cortex. Statistics and geometry. Springer, Berlin Heidelberg New York

Breakspear M, Terry J, Friston KJ (2003) Modulation of excitatory synaptic coupling facilitates synchronization and complex dynamics in a nonlinear model of neuronal dynamics Network: Computation in Neural Systems 14, 703–732

Breakspear M (2004) "Dynamic" connectivity in neural systems: Theoretical and empirical considerations. Neuroinformatics 4,1–23.

Breakspear M, Stam KJ (2005) Dynamics of a neural system with a multiscale architecture. Phil. Trans. R. Soc. B 360, 1051–1074.

Breakspear M, Roberts JA, Terry JR, Rodrigues S, Robinson PA (2006) A unifying explanation of generalized seizures via the bifurcation analysis of a dynamical brain model. Cerebral Cortex, doi:10.1093/cercor/bhj072.

Bressler SL (1990) The gamma wave: a cortical information carrier? Trends in Neuroscience 13(5), 161–162

Bressler SL (1995) Large-scale cortical networks and cognition. Brain Res. Rev. 20, 288–304

Bressler SL (2002) Understanding cognition through large-scale cortical networks. Curr. Dir. Psych. Sci. 11, 58–61

Bressler SL (2003) Cortical coordination dynamics and the disorganization syndrome in schizophrenia. Neuropsychpharmacology 28, 535–539

Bressler SL, Kelso JAS (2001) Cortical coordination dynamics and cognition. Trends in Cog. Sci. 5, 26–36

Bressler SL, Tognoli E (2006) International Journal of Psychophysiology (in press)

Bressler SL, McIntosh AR (2007) The role of neural context in large-scale neurocognitive network operations. This Volume.

Brunel N, Hakim V (1999) Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. Neural Comput. 11, 1621–1671

Bullmore ET, Rabe-Hesketh S, Morris RG, Williams SC, Gregory L, Gray JA, Brammer MJ (1996) Functional magnetic resonance image analysis of a large-scale neurocognitive network. Neuroimage 4, 16–33

Campbell SA (2007) Time Delays in Neural Systems. This Volume.

Cessac B, Samuelides M (2006) From Neuron to Neural Network Dynamics. To appear in Dynamical Neural Network. Models and Applications to Neural Computation. Springer Berlin Heidelberg New York

Chu PH, Milton JG, Cowan JD (1994) Connectivity and the dynamics of integrate-and-fire neural networks. Int. J. Bifur. Chaos 4, 237–217

Churchland P, Sejnowski TJ (1992) The Compuational Brain. MIT Press. New York.

Collet P, Eckmann J (1980) Iterated maps on the interval as dynamical systems, Birkhauser.

Coombes S, Lord GJ, Owen MR (2003) Waves and bumps in neuronal networks with axo-dendritic synaptic interactions. Physica D, 178, 219–241

Coombes S (2005) Waves, bumps, and patterns in neural field theories, Biological Cybernetics 93, 91–108

Crick F, Koch C ((1990) Towards a neurobiological theory of consciousness. Seminars in the Neurosciences 2, 263–275

Cvitanovic P (1984) Universality in chaos. Adam Hilger: Bristol.

Dayan P, Abbott LF (2001) Theoretical Neuroscience. MIT Press Cambridge, Massachusetts

De Monte S, d'Ovidio F, Mosekilde E (2003) Coherent regimes of globally coupled dynamical systems. Phys. Rev. Let. 90, 054102

Dhamala M, Jirsa VK, Ding M (2004) Transitions to synchrony in coupled bursting neurons. Phys. Rev. Lett. 92: 028101.
Eckmann J, Ruelle D (1985) Ergodic theory of chaos and strange attractors. Reviews of Modern Physics, 57: 617–656
Erlhagen W, Schöner G (2002) Dynamic field theory of movement preparation.. Psychological Review 109, 545–572
Ermentrout B (1998) Neural networks as spatio-temporal pattern-forming systems. Rep. Prog. Phys. 61: 353–430.
Feigenbaum MJ (1978) Quantitative universality for a class of nonlinear transformations. J. Stat. Phys., 19: 25–52.
Ferree TC, Nunez PL (2007) Primer on electroencephalography for functional connectivity. This volume
FitzHugh R (1961) Impulses and physiological states in theoretical models of nerve membrane. Biophys. J. 1, 445–466
Frank TD, Daffertshofer A, Beek PJ, Haken H (1999) Impacts of noise on a field theoretical model of the human brain. Physica D, 127, 233–249
Frank TD, Daffertshofer A, Peper CE, Beek PJ, Haken H (2000) Towards a comprehensive theory of brain activity: coupled oscillator systems under external forces. Physica D, 144, 62–86
Freeman WJ (1975) Mass action in the nervous system. Academic Press New York
Freeman WJ, Skarda CA (1985) Spatial EEG patterns, nonlinear dynamics and perception: the neo-Sheringtonian view. Brain Res Rev10, 147–175
Freeman WJ (1987) Simulation of chaotic EEG patterns with a dynamic model of the olfactory system. Biological Cybernetics, 56, 139–150
Freeman WJ (1992) Tutorial on neurobiology: From single neurons to brain chaos. Inter. Journ. Bif. Chaos 2, 451–482
Fuchs A, Jirsa VK, Kelso JA (2000) Theory of the relation between human brain activity (MEG) and hand movements. Neuroimage 11(5), 359–369
Fujisaka H, Yamada T (1983) Stability theory of synchronized motion in coupled-oscillator system. Progress in Theoretical Physics, 69, 32–47.
Fusi S, Drew PJ, Abbott LF (2005). Cascade models of synaptically stored memories. Neuron. 45, 599–611.
Gardiner CW (2004) Handbook of Stochastic Methods. Springer Berlin Heidelberg New York
Gerstner W (2000) Population Dynamics of Spiking Neurons: Fast Transients, Asynchronous States, and Locking. Neural Computation 12, 43–89
Gerstner W, Kistler WM (2002) Spiking neuron models: Single neurons, populations, plasticity. Cambridge University Press
Gray CM, Singer W (1989) Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. Proc. Nat. Acad. Sci. 86, 1698–1702
Grebogi C, Ott E, Yorke JA (1982) Chaotic atractors in crisis. Phys. Rev. Lett. 48: 1507–1510.
Griffith JS (1963) A field theory of neural nets: I. Derivation of field equations. Bull. Math. Biophys. 25, 111–120
Griffith JS (1965) A field theory of neural nets: II. Properties of the field equations. Bull. Math. Biophys. 27, 187–195
Gröbler T, Barna Gy, Érdi P (1988) Statistical model of the hippocampal CA3 region II. The population frame work: model of rhythmic activity in the CA3 slice. Biol. Cybern. 79, 309–321

Grossberg S (1988) Nonlinear Neural Networks: Principles, Mechanisms, and Architectures. Neural Networks 1, 17–61

Guckenheimer, J. (1987) Limit sets of S-unimodal maps with zero entropy, Communications in Mathematical Physics, 110: 655–659.

Guevara MG (2003). Dynamics of excitable cells. In: Nonlinear Dynamics in Physiology and Medicine (A Beuter, L Glass, MC Mackey and MS Titcombe, eds). Springer-Verlag, New York, pp. 87–121.

Haken H (1983) Synergetics. An introduction. 3rd edition. Springer Berlin, Heidelberg, New York

Haken H (1999) Information and Self-Organization. 2nd edition. Springer Berlin, Heidelberg, New York

Hindmarsh JL, Rose RM (1984) A model of neuronal bursting using three coupled first order differential equations. Proc. R. Soc. London, Ser. B 221, 87

Hodgkin AL, Huxley AF (1952) A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve/Journal of Physiology, 117: 500–544

Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc. Nat. Acad. Sci. 79, 2554–2558

Hopfield JJ (1984) Neurons with graded response have collective computational properties like those of two-state neurons. Proc. Nat. Acad. Sci. 81, 3088–3092

Horwitz B, Friston KJ, Taylor JG (2000) Neural modeling and functional brain imaging: an overview. Neural Networks 13, 829–846

Izhikevich E (2005) Dynamical systems in neuroscience: The geometry of excitability and bursting. MIT Press.

Jirsa VK, Haken H (1996) Field theory of electromagnetic brain activity. Physical Review Letters, 77: 960–963.

Jirsa VK, Haken H (1997) A derivation of a macroscopic field theory of the brain from the quasi-microscopic neural dynamics. Physica D 99: 503–526.

Jirsa VK, Fuchs A, Kelso JAS (1998) Connecting cortical and behavioral dynamics: bimanual coordination. Neural Computation 10, 2019–2045

Jirsa VK, Kelso JAS. (2000) Spatiotemporal pattern formation in continuous systems with heterogeneous connection topologies. Phys. Rev. E 62, 6, 8462–8465

Jirsa VK, Jantzen KJ, Fuchs A, Kelso JAS (2002) Spatiotemporal forward solution of the EEG and MEG using network modeling. IEEE Transactions on Medical Imaging, 21, 5, 493–504

Jirsa VK (2004) Connectivity and dynamics of neural information processing. Neuroinformatics 2 (2), 183–204

Jirsa VK (2004b) Information processing in brain and behavior displayed in large-scale scalp topographies such as EEG and MEG. Inter. J. Bif. Chaos 14(2), 679–692

Kaneko K (1997) Dominance of Milnor attractors and noise-induced selection in a multiattractor system. Physical Review Letters, 78: 2736–2739

Kiss T, Érdi P (2002) Mesoscopic Neurodynamics. BioSystems 64, 119–126

Koch C (1999) Biophysics of Computation. Information processing in single neurons. Oxford University Press

Larter R, Speelman B and Worth R M (1999) A coupled ordinary differential equation lattice model for the simulation of epileptic seizures. *Chaos*, **9**: 795–804.

Liley DTJ, Cadusch PJ, Dafilis MP (2002) A spatially continuous mean field theory of electrocortical activity. Network-Computation in Neural Systems, 13, 67–113.
Lopes da Silva FH, Hoeks A, Smits H, Zetterberg LH (1974) Model of brain rhythmic activity: the alpha-rhythm of the thalamus. Kybernetik 15,27–37
Lopes da Silva FH, Blanes W, Kalitzin S, Parra J, Suffczynski P, Velis DN (2003) Dynamical diseases of brain systems: different routes to epileptic seizures Trans. Biomed. Eng. 50: 540–548.
Lorenz, E (1963) Deterministic nonperiodic flow. Journal of Atmospheric Science, 20: 130–141.
Maistrenko Y, Maistrenko V, Popovich A, Mosekilde E (1998) Transverse instability and riddled basins in a system of two coupled logistic maps. Physical Review E, 57: 2713–2724.
Mallot HA, Brittinger R (1989) Towards a network theory of cortical areas. In: Cotteril RMJ. (ed) Models of brain function. Cambridge University Press, 175–189
Mallot HA, Giannakopoulos F (1996) population networks: a large-scale framework for modelling cortical neural networks. Biological Cybernatics 75, 441–452
McCormick DA, Bal T (1997) Sleep and Arousal: Thalamocortical mechanisms. Ann. Rev. Neuroscience, 20: 185–215
McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5:115–33.
McIntosh AR (2000) Towards a network theory of cognition. Neural Netw. 13, 861–876
Mesulam MM (1998) From sensation to cognition. Ann. Neurol. 28, 597–613
Milton JG (1996) Dynamics of small neural populations. American Mathematical Society
Milton JG, Chkhenkeli SA, Towle VL (2007) Brain Connectivity and the Spread of Epileptic Seizures. This volume.
Milnor J (1985) On the concept of attractor. Communications in Mathematical Physics, 99: 177–195.
Morris C, Lecar H (1981) Voltage oscillations in the barnacle giant muscle fiber. Biophysics J., 35: 193–213
Mountcastle VB (1978) An organizing principle for cerebral function: the unit module and the distributed system. In: Edelman GM, Mountcastle VB (Eds) The Mindful Brain. MIT Press, Cambridge MA
Mountcastle VB (1998) Perceptual Neuroscience: the cerebral cortex. Harvard University Press, Cambridge MA
Nagumo J, Arimoto S, Yoshizawa S (1962) An active pulse transmission line
simulating nerve axon. Proc IRE. 50: 2061–2070.
Nunez PL (1974) The brain wave equation: a model for the EEG. Mathematical Biosciences 21: 279–297.
Nunez PL (1995) Neocortical dynamics and human EEG rhythms, Oxford University Press
Nykamp DQ, Tranchina D (2001) A population density approach that facilitates large-scale modeling of neural networks: Extension to slow inhibitory synapses. Neural Computation 13, 511–546
Nykamp DQ, Tranchina D (2000) A population density approach that facilitates large-scale modeling of neural networks: analysis and an application to orientation tuning. Journal of Computational Neuroscience 8, 19–50

Ratliff F, Knight BW, Graham N (1969) On tuning and amplification by lateral inhibition. PNAS 3: 733–740.
Robinson PA, Rennie CJ, Wright JJ (1997) Propagation and stability of waves of electrical activity in the cerebral cortex. Physical Review E, 56: 826–840.
Robinson PA, Rennie CJ, Wright JJ, Bahramali H, Gordon E, Rowe DL (2001) Prediction of electroencephalographic spectra from neurophysiology. Physical Review E 63, 021903
Robinson PA, Rennie CJ, Rowe DL (2002) Dynamics of large-scale brain activity in normal arousal states and epileptic seizures. Physical Review E 65, 041924.
Robinson PA, Rennie CJ, Rowe DL, O'Connor SC (2004) Hum Brain Mapp 25, 53–72
Robinson PA, Rennie CJ, Rowe DL, O'Connor SC, Gordon E (2005) Phil. Trans. Roy. Soc. Ser. B 360, 1043
Rodriguez, S, Terry JR, Breakspear M (2006) On the genesis of spike-wave activity in a mean-field model of human corticothalamic dynamics. Physics Letters A 355, 352–357
Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain, Psychol. Rev. 65: 386–408.
Rulkov N, Sushchik M, Tsimring L, Abarbenel H (1995) Generalized synchronization of chaos in unidirectionally coupled chaotic systems. *Physical Review E*, **51**: 980–994.
Sejnowski TJ, Rosenberg CR (1987) Parallel networks that learn to pronounce English. Complex Systems, 1: 145–168.
Smalheiser NR (2000) Walter Pitts. Perspectives in Biology and Medicine. 43: 217–226.
Sporns O (2003) Complex Neural Dynamics. In: Coordination Dynamics: Issues and Trends. Jirsa VK & Kelso JAS (eds.) Springer Berlin
Sporns O, Tononi G (2002) Classes of Network connectivity and dynamics. Complexity 7, 28–38
Sporns O, Tononi G (2007) Structural determinants of functional brain dynamics. This Volume.
Strogatz SH (1994) Nonlinear dynamics and Chaos. Addison-Wesley: Reading, MA.
Steyn-Ross ML, Steyn-Ross DA, Sleigh JW, Liley DTJ (1999) Theoretical electroencephalogram stationary spectrum for a white-noise-driven cortex: Evidence for a general anesthetic-induced phase transition. Phys. Rev. E 60, 7299–7311
Szentagothai J (1975) The 'module-concept' in cerebral cortex architecture. Brain Res. 95, 476–496
Tagamets MA, Horwitz B (1998) Integrating electrophysiological and anatomical experimental data to create a large-scale model that simulates a delayed match-to-sample human brain imaging study. Cereb. Cortex 8, 310–320
Treisman A (1996) The binding problem. Curr. Pin. Neurobiol. 6, 171–178
van Rotterdam A, Lopes da Silva FH, van den Ende J, Viergever MA, Hermans AJ (1982) A model of the spatio- temporal characteristics of the alpha rhythm. Bulletin of Mathematical Biology. 44: 283–305.
Valdes PA, Jimenez JC, Riera J, Biscay R, Ozaki T (1999) Nonlinear EEG analysis on a neural mass model. Biol. Cybern. 81, 415–424
Ventriglia F (1974) Kinetic approach to neural system. Bull. Math. Biol. 36, 535–544
Ventriglia F (1978) Propagation of excitation in a model of neural system. Biol. Cybern. 30, 75–79

Von Stein A, Rappelsberger P, Sarnthein J, Petsche H (1999) Synchronization between temporal and parietal cortex during multimodal object processing in man. Cereb. Cortex 9, 137–150

Wilson HR (1973) Cooperative phenomena in a homogenous cortical tissue model. In: Haken H. (ed.) Synergetics – Cooperative Phenomena in Multi-compartment Systems. B. G. Teubner, Stuttgart

Wilson HR and Cowan JD (1972) Excitatory and inhibitory interactions in localized populations of model neurons. Biophys. J. 12, 1–23.

Wilson HR, Cowan JD (1973) A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. Kybernetic 13, 55–80

Wright JJ, Liley DTJ (1996) Dynamics of the brain at global and microscopic scales: Neural networks and the EEG. Behav. Brain. Sci. 19, 285

Time Delays in Neural Systems

Sue Ann Campbell

Department of Applied Mathematics, University of Waterloo, Waterloo ON N2L 3G1 Canada sacampbell@uwaterloo.ca
Centre for Nonlinear Dynamics in Physiology and Medicine, McGill University, Montréal Québec H3C 3G7 Canada

1 Introduction

In this chapter I will give an overview of the role of time delays in understanding neural systems. The main focus will be on models of neural systems in terms of delay differential equations. Later in this section, I will discuss how such models arise. The goal of the chapter is two-fold: (1) to give the reader an introduction and guide to some methods available for understanding the dynamics of delay differential equations and (2) to review some of the literature documenting how including time delays in neural models can have a profound effect on the behaviour of those models.

1.1 Modelling Delay in Neural Systems

To begin, I will formulate a general model for a network of neurons and then determine how delays may occur in this model. Consider a network of n neurons modelled by the equations

$$\dot{\mathbf{x}}_i(t) = \mathbf{F}_i(\mathbf{x}_i(t)) + \sum_{j=1}^{n} \mathbf{f}_{ij}(\mathbf{x}_i(t), \mathbf{x}_j(t)), \ i = 1, \ldots, n \ . \tag{1}$$

The variable $\mathbf{x}_i$ represents all the variables describing the physical state of the cell body of the i^{th} neuron in the network. For example, in the standard Hodgkin-Huxley model, it would represent the membrane voltage and gating variables: $\mathbf{x}_i = (V_i, m_i, n_i, h_i)$. The function $\mathbf{F}_i$ represents the intrinsic dynamics of the i^{th} neuron and the function $\mathbf{f}_{ij}$, often called the coupling function, represents the input to the i^{th} neuron from the j^{th} neuron. In neural models, the coupling is usually through the voltage, V_i, only, so $\mathbf{f}_{ij} = [f_{ij}, 0, 0, \ldots, 0]^T$. I will primarily consider this case in the rest of the chapter.

If the j^{th} neuron is connected to the i^{th} via a chemical synapse, then the coupling function is given by

$$f_{ij}(\mathbf{x}_i(t), \mathbf{x}_j(t)) = c_{ij}\, h_{ij}^{\text{pre}}(\mathbf{x}_j(t))\, h_{ij}^{\text{post}}(\mathbf{x}_i(t)) \,. \tag{2}$$

This is called **synaptic** coupling. Here h_{ij}^{pre} is a sigmoidal function, usually chosen to have maximum value 1, so that c_{ij} represents the maximum coupling strength (synaptic conductance). h_{ij}^{post} is typically a linear function (e.g. for Hodgkin Huxley models, $h_{ij}^{\text{post}}(\mathbf{x}_i(t)) = V_i(t) - K_{ij}$, where K_{ij} is a constant). Some models set $\mathbf{h}_{ij}^{\text{post}} = 1$, eliminating the dependence on the post-synaptic neuron, in which case this coupling is called **sigmoidal**.

If the neurons are connected via a gap junction, then the coupling function is

$$\mathbf{f}_{ij}(\mathbf{x}_i(t), \mathbf{x}_j(t)) = C_{ij}(\mathbf{x}_i(t) - \mathbf{x}_j(t)) \,, \tag{3}$$

where C_{ij} is the matrix of coupling coefficients. This is called **gap junctional**, **electrical** or **diffusive** coupling. For most neural models only the $(1,1)$ element of C_{ij} is non zero.

There are several sources of delay in a neural system. Consider first the delay due to propagation of action potentials along the axon. In the model above, when an action potential is generated in the cell body of neuron j, it is immediately felt by all other neurons to which it is connected. However, in reality, the action potential must travel along the axon of neuron j to the synapse or gap junction. Conduction velocities can range from the order of 1 m/sec along unmyelinated axons to more than 100 m/sec along myelinated axons (Desmedt and Cheron, 1980; Shepherd, 1994). This can lead to significant time delays in certain brain structures. There are several ways to incorporate this into the model, such as including spatial dependence of the variables or multiple compartments representing different parts of the neuron (Koch, 1990). However, if we are primarily interested in the effect of the action potential when it reaches the end of the axon (will it cause an action potential in another neuron?), then a simpler approach is to include a time delay in the coupling term. In this case the general coupling term becomes

$$f_{ij}(\mathbf{x}_i(t), \mathbf{x}_j(t - \tau_{ij})) \tag{4}$$

where $\tau_{ij} > 0$ represents the time taken for the action potential to propagate along the axon connecting neuron j (the pre-synaptic neuron) to neuron i (the post-synaptic neuron).

The above assumes that the axon of neuron j connects on or close to the cell body of neuron i. Some cells may have synapses or gap junctions on dendrites far from the cell body. In this case, there can also be a delay associated with propagation of the action potential along the dendrite. This will introduce an additional time delay, viz.,

$$f_{ij}(\mathbf{x}_i(t - \tau_{ij}^d), \mathbf{x}_j(t - \tau_{ij}^d - \tau_{ij})) \tag{5}$$

where τ_{ij} and τ_{ij}^d represent the time delays due to the propagation of the action potential along the axon and dendrite, respectively.

Another delay can occur in the transmission of the signal across the synapse. That is, once the action potential from neuron j reaches the synapse, there is some time before an action potential is initiated in neuron i. A common way to model this is to augment the model equations above by equations modelling the chemical kinetics of the synapse (Keener and Sneyd, 1998; Koch, 1999). Alternatively, this can be incorporated into (4) or (5) just by increasing the delay τ_{ij}. I will focus on the latter approach, but in Sect. 3 will review some literature that indicates the qualitative effect on the dynamics can be quite similar using both approaches. Clearly, the latter approach will yield a simpler model if one also wants to include the effect of axonal delay.

Equations (4) and (5) assume that the time delays are fixed. In reality, the delay will likely vary slightly each time an action potential is propagated from neuron j to neuron i. This may be incorporated into the model putting time dependence into the delay: $\tau_{ij}(t)$. Many of the methods outlined in Sect. 2 may be extended to this case, by assuming the delay satisfies some constraints $0 \le \tau_{ij}(t) \le \bar{\tau}_{ij}$. Alternatively, one might consider adding some noise to the delay, which would lead to a stochastic delay differential equation model. Unfortunately, there is very little theory available for such equations.

An alternative approach is to incorporate a distribution of delays, representing the situation where the delay occurs in some range of values with some associated probability distribution. In this case, coupling term (4) becomes

$$\int_0^\infty f_{ij}(\mathbf{x}_i(t), \mathbf{x}_j(t-\sigma)) g_{ij}(\sigma)\, d\sigma \ , \tag{6}$$

and similarly for (5). The function g_{ij} is called the kernel of the distribution and represents the probability density function of the time delay. Since g_{ij} is a pdf it is normalized so that $\int_0^\infty g_{ij}(\sigma)\, d\sigma = 1$. Although distributions of delays are not commonly used in neural network models, they have been extensively used in models from population biology (Cushing, 1977; MacDonald, 1978). In this literature, the most commonly used distributions are the uniform distribution:

$$g_{ij}(\sigma) = \begin{cases} 0 & 0 \le \sigma < \tau_{ij}^{\min} \\ \frac{1}{\delta} & \tau_{ij}^{\min} \le \sigma \le \tau_{ij}^{\min} + \delta \\ 0 & \tau_{ij}^{\min} < \sigma \end{cases} \ , \tag{7}$$

and the gamma distribution:

$$g_{ij}(\sigma) = \begin{cases} 0 & 0 \le \sigma < \tau_{ij}^{\min} \\ \frac{a^m}{\Gamma(m)} (\sigma - \tau_{ij}^{\min})^{m-1} e^{-a(\sigma - \tau_{ij}^{\min})} & \tau_{ij}^{\min} \le \sigma \end{cases} \ , \tag{8}$$

where $a, m \ge 0$ are parameters. Γ is the gamma function defined by $\Gamma(0) = 1$ and $\Gamma(m+1) = m\Gamma(m)$. Both these distributions can be shown to approach a Dirac distribution in the appropriate limits ($\delta \to 0$ for the uniform distribution and $m \to \infty$ for the gamma distribution), which leads to a discrete delay in the coupling term. It is usual in the population biology literature

(Cushing, 1977; MacDonald, 1978) to take $\tau_{ij}^{min} = 0$. In this case model with a gamma distribution can be shown to be equivalent to a system of m ordinary differential equations, which is amenable to the analysis techniques described elsewhere in this volume (Breakspear and Jirsa, 2006). However, as pointed out by Bernard et al. (2001), it makes more biological sense to take $\tau_{\min} > 0$, since the probability of having zero delay is effectively zero in most applications. In this case, the model with a gamma distribution is equivalent to a system of $m-1$ ordinary differential equations and one discrete delay differential equation.

In the next section I will review some tools for analyzing delay differential equations. To make the theory concrete, we will apply it to a particular example. Consider the following representation of the Fitzhugh-Nagumo model for a neuron (Fitzhugh, 1960; Nagumo et al., 1962)

$$\begin{aligned} \dot{v}(t) &= -v^3 + (a+1)v^2 - av - w + I \ , \\ \dot{w}(t) &= bv - \gamma w \ . \end{aligned} \tag{9}$$

Assume that the parameters are such that there is just one equilibrium point $(\bar{v}, \bar{w})$ of this equation, where $\bar{v}, \bar{w}$ satisfy

$$\bar{v}^3 - (a+1)\bar{v}^2 + (a + \frac{b}{\gamma})\bar{v} + I = 0 \ , \tag{10}$$

$$\bar{w} = \frac{b}{\gamma}\bar{v} \ . \tag{11}$$

I shall consider the situation when two of these neurons are joined with delayed sigmoidal coupling in the following way

$$\begin{aligned} \dot{v}_1(t) &= -v_1^3 + (a+1)v_1^2 - av_1 - w_1 + I + c\tanh(v_2(t-\tau) - \bar{v}) \\ \dot{w}_1(t) &= bv_1 - \gamma w_1 \\ \dot{v}_2(t) &= -v_2^3 + (a+1)v_2^2 - av_2 - w_2 + I + c\tanh(v_1(t-\tau) - \bar{v}) \\ \dot{w}_2(t) &= bv_2 - \gamma w_2 \end{aligned} \tag{12}$$

This setup is due to Burić et al. (2005). I will sometimes write (12) in the condensed form

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}(t), \mathbf{x}(t-\tau)) \ , \tag{13}$$

where $\mathbf{x} = (v_1, w_1, v_2, w_2)$.

I will focus on equations with a single discrete delay. The approach is similar for multiple delays, the analysis just becomes more complicated. We will discuss some of the differences that arise for distributed delays in the final section.

There is a very large literature on the effect of time delays on artificial neural networks (ANNs). An example of such a network is the additive (also called Hopfield) neural network with delays. This is usually written in the form

$$\dot{x}_i(t) = -k_i x_i(t) + \sum_{j=1}^{n} f_{ij}(x_j(t-\tau_{ij}) \ .$$

I will not attempt to review all the material related to such equations, but will try to highlight those results I feel may have implications for biological neural networks. In particular, networks of the following form

$$\dot{x}_i(t) = -k_i x_i(t) + f_{ii}(x(t-\tau_1)) + \sum_{j\neq i}^{n} f_{ij}(x_j(t-\tau_2)) \,, \tag{14}$$

have some parallels with biological neural networks, since the uncoupled units may behave as type II oscillators (Campbell et al., 2005).

2 Tools for Analysis

The main tools for studying the behaviour of delay differential equations are extensions of those for ordinary differential equations which are discussed elsewhere in this volume (Breakspear and Jirsa, 2006). Some familiarity with these tools will be helpful in reading this section.

To improve the flow of the text, I will not give references for all the standard results for delay differential equations that I use. For more information on these, I refer the reader to the fairly accessible books of Kolmanovskii and Nosov (1986) and Stépán (1989) which cover the results of this section or the books of Hale and Verduyn Lunel (1993) and Diekmann et al. (1995) which give complete, although not so accessible, accounts of the theory of delay differential equations.

To begin our discussion, consider the types of solutions which occur most often in neural systems. These are equilibrium solutions ($\mathbf{x}(t) = \bar{\mathbf{x}}$, for some constant $\bar{\mathbf{x}}$) and periodic solutions ($\mathbf{x}(t) = \mathbf{x}(t+T)$ for some $T > 0$). The fundamental questions we would like to answer in order to understand the behaviour of a model with time delays are the following

1. *What equilibrium solutions occur in the system?*
2. *What periodic solutions occur in the system?*
3. *Are these stable or unstable? That is, do we expect to observe them in experiments and numerical simulations?*
4. *How do the answers to these questions change as parameters are varied?*

Question 1 is easily answered: the equilibrium solutions of a system with time delays are the same as those of the corresponding system with zero delay. Thus for (13) these correspond to constant vectors $\bar{\mathbf{x}}$ such that $\mathbf{f}(\bar{\mathbf{x}}, \bar{\mathbf{x}}) = 0$.

Example. For system (12) the equilibrium points are given by $(v_1, w_1, v_2, w_2) = (\bar{v}_1, \bar{w}_1, \bar{v}_2, \bar{w}_2)$ where $\bar{v}_j, \bar{w}_j$ are constants, found by solving the equations

$$\begin{aligned} 0 &= -\bar{v}_1^3 + (a+1)\bar{v}_1^2 - a\bar{v}_1 - \bar{w}_1 + I + c\tanh(\bar{v}_2 - \bar{v}) \\ 0 &= b\bar{v}_1 - \gamma\bar{w}_1 \\ 0 &= -\bar{v}_2^3 + (a+1)\bar{v}_2^2 - a\bar{v}_2 - \bar{w}_2 + I + c\tanh(\bar{v}_1 - \bar{v})) \\ 0 &= b\bar{v}_2 - \gamma\bar{w}_2 \end{aligned} \tag{15}$$

It is easy to check that one solution of these equations is $(\bar{v}_1, \bar{w}_1, \bar{v}_2, \bar{w}_2) = (\bar{v}, \bar{w}, \bar{v}, \bar{w})$, where $\bar{v}, \bar{w}$ are given by (10)–(11). I will focus on this solution in later discussions of this example.

Question 2 is difficult to answer analytically with any completeness. A partial answer can be obtained by determining the bifurcations that occur in the system which lead to the creation of periodic solutions. More detail can be found in subsection 2.2. This question can also be addressed through the use of numerical tools, which are discussed in subsection 2.5.

For equilibrium solutions, question 3 can be addressed via linear stability analysis (see subsection 2.1) and via Lyapunov theory (see subsection 2.3). For periodic solutions this question generally must be answered using numerical tools, as discussed in subsection 2.5.

Answering question 4 is the main goal of bifurcation theory. Analytical methods for studying bifurcations will be discussed in subsection 2.2 and numerical methods in subsection 2.5.

2.1 Linear Stability

One way to study the stability of an equilibrium solution is through linearization. This is constructed in a similar way as for ordinary differential equations. The **linearization** of (13) about $\bar{\mathbf{x}}$ is given by

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{x}(t-\tau) \tag{16}$$

where A is a the Jacobian matrix of $\mathbf{f}(y,z)$ with respect to y, i.e. the matrix with entries $a_{ij} = \frac{\partial f_i}{\partial y_j}$, and B is the Jacobian matrix of $\mathbf{f}(y,z)$ with respect to z. If the system has multiple delays, then there will be a term in the linearization corresponding to each delay.

It can be shown that, under the right conditions, (16) describes the behaviour of solutions close to $\bar{\mathbf{x}}$. This will in turn determine the stability of $\bar{\mathbf{x}}$. To study this behaviour, we assume that there are solutions of (16) of the form $\mathbf{x}(t) = e^{\lambda t}\mathbf{k}$ where λ is a complex number and $\mathbf{k}$ is an n-vector of complex numbers, to be determined. Substituting this into (16) we obtain

$$\left[-\lambda I + A + Be^{-\lambda\tau}\right]\mathbf{k} = 0\,. \tag{17}$$

For solutions with $\mathbf{k} \neq 0$ to exist, we require

$$\det[-\lambda I + A + Be^{-\lambda\tau}] = 0\,. \tag{18}$$

If (13) is an n-dimensional system, then (18) can be written in the form

$$\Delta(\lambda) = \lambda^n + \lambda^{n-1}(\delta_{n-1,0} + \delta_{n-1,1}e^{-\lambda\tau}) + \cdots + \lambda\sum_{j=0}^{n-1}\delta_{1,j}e^{-j\lambda\tau} + \sum_{j=0}^{n}\delta_{0,j}e^{-j\lambda\tau} = 0\,, \tag{19}$$

where the $\delta_{i,j}$ depend on the elements of the matrices A and B.

Equation (19) is called the **characteristic equation** of the linearization of (13) about $\bar{\mathbf{x}}$. Any complex number λ which satisfies (19) will give rise to a solution of (16) ($\mathbf{k}$ is found by solving (17) with the particular value of λ substituted in). In practice, we are mostly concerned with the λ values for the reasons outlined below.

Example. For our coupled Fitzhugh-Nagumo model (12) the linearization about the equilibrium point $(\bar{v}, \bar{w}, \bar{v}, \bar{w})$ is given by (16) where

$$A = \begin{bmatrix} \alpha & -1 & 0 & 0 \\ b & -\gamma & 0 & 0 \\ 0 & 0 & \alpha & -1 \\ 0 & 0 & b & -\gamma \end{bmatrix} \text{ with } \alpha = -3\bar{v}^2 + 2(a+1)\bar{v} - a, \text{ and } B = \begin{bmatrix} 0 & 0 & c & 0 \\ 0 & 0 & 0 & 0 \\ c & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} .$$

Note that α depends on all the intrinsic neuron parameters (a, b, γ, I), since $\bar{v}$ is a solution of (10). Putting A, B into (18) shows that the characteristic equation for this example is

$$\Delta_+(\lambda)\Delta_-(\lambda) = 0 \tag{20}$$

where

$$\Delta_\pm(\lambda) = (\lambda + \gamma)(\lambda - \alpha \pm ce^{-\lambda\tau}) + b \, .$$

Fact: If all the roots of the characteristic equation of the linearization of (13) about $\bar{\mathbf{x}}$ have negative real part, then $\bar{\mathbf{x}}$ is asymptotically stable, i.e., all solutions which start sufficiently near to $\bar{\mathbf{x}}$ will tend toward it as t increases.

Fact: If at least one root of the characteristic equation of the linearization of (13) about $\bar{\mathbf{x}}$ has positive real part, then $\bar{\mathbf{x}}$ is unstable, i.e., some solutions which start near to $\bar{\mathbf{x}}$ will tend away from it as t increases.

So we see that to determine the stability of an equilibrium point we need to determine the roots, λ of the characteristic (19). These are often called the **eigenvalues** of the equilibrium point. For ordinary differential equations, the characteristic equation is a polynomial in λ and hence there are a finite number of solutions all of which may be calculated or numerically approximated. For delay differential equations, however, the presence of the $e^{-\lambda\tau}$ terms means that there are an *infinite* number of solutions of the characteristic equation. This means we must rely on other methods to determine whether an equilibrium point is stable. Several methods are outlined in the book of Kolmanovskii and Nosov (1986), here we will focus on a particular one which relies on the following result.

Fact: The zeros of $\Delta(\lambda)$ depend continuously on τ and the $\delta_{i,j}$, and hence on the elements of A and B. Thus as any of these parameters is varied, the number of zeros of $\Delta(\lambda)$ with positive real part can only change if a root passes through the imaginary axis.

The most common way of using this fact in coupled neural systems, is outlined in the following procedure.

1. Set the delay, τ, equal to zero. This will change the delay differential equation into an ordinary differential equation with the same equilibrium points as the delay differential equation.
2. Determine the stability of an equilibrium point for the ODE system, i.e. determine the number of eigenvalues with positive real parts.
3. Determine the critical values of the delay, $\tau_1^c < \tau_2^c < \cdots$ for which the characteristic (19) has eigenvalues with zero real parts. These are the values of the delay where the stability of the equilibrium point may change.
4. Calculate the rate of change of the real part of an eigenvalue with respect to τ when τ is equal to one of the critical values found in the previous step, i.e., calculate

$$\left.\frac{dRe(\lambda)}{d\tau}\right|_{\tau=\tau_k^c} = -\,Re\left(\frac{\partial\Delta}{\partial\tau}\Big/\frac{\partial\Delta}{\partial\lambda}\right)\Bigg|_{\tau=\tau_k^c} .$$

   If $\frac{dRe(\lambda)}{d\tau} > 0$, then the number of roots with positive real parts is increasing, if it is negative, then the number of roots is decreasing.
5. Due to the fact above, the number of roots of the characteristic equation with positive real part will be constant for $0 \le \tau < \tau_1$ and equal to the number found in step 2. For each subsequent interval, $\tau_k < \tau < \tau_{k+1}$, the number can be determined from the number in the previous interval $\tau_{k-1} < \tau < \tau_k$, the number of roots with zero real part at τ_k and the rate of change calculated in step 4.

Example. Consider our coupled Fitzhugh-Nagumo model (12). We will follow the procedure outlined above.

1. When $\tau = 0$ the factors of the characteristic (20) become

$$\Delta_\pm = \lambda^2 + \lambda(\gamma - \alpha \pm c) + \gamma(-\alpha \pm c) + b\ .$$

2. By analyzing the roots of this equation, it can be shown that if $\gamma^2 < b$ the trivial solution is stable for $|c| < \gamma - \alpha \stackrel{def}{=} c_H$, and for c outside this region the equilibrium point has two complex conjugate eigenvalues with positive real part, i.e. it is unstable. (In fact the two points $c = \pm c_H$ are Hopf bifurcation points for the system with zero delay.)
3. To find the parameter values where the characteristic (20) has eigenvalues with zero real part, we substitute $\lambda = i\omega$ into (20) and separate into real and imaginary parts. This yields

$$\begin{aligned} -\alpha + \frac{b\gamma}{\gamma^2+\omega^2} \pm c\cos\omega\tau &= 0 \\ \omega\left(1 - \frac{b}{\gamma^2+\omega^2}\right) \mp c\sin\omega\tau &= 0\ . \end{aligned}$$

Note that we choose the + in the first equation and − in the second for the parameter values for Δ_+ to have a pair of complex conjugate roots and the opposite for Δ_-. Some rearrangement of these equations gives

$$(b\gamma - \alpha(\gamma^2 + \omega^2))^2 + \omega^2(\gamma^2 + \omega^2 - b)^2 - c^2(\gamma^2 + \omega^2)^2 = 0 \tag{21}$$

and

$$\tan \omega\tau = \frac{\omega(\gamma^2 + \omega^2 - b)}{b\gamma - \alpha(\gamma^2 + \omega^2)} \,. \tag{22}$$

Thus, for given values of the parameters a, b, γ, I (which determine α) and c one can find ω from the first equation and the corresponding τ values from the second equation. Alternatively, we can think of these two equations as defining the coupling parameters τ and c in terms of the intrinsic neuron parameters and ω. Then these equations define curves in the c, τ parameter plane. These curves are shown in Fig. 1 for a specific set of intrinsic parameter values. There are multiple curves because tan is a periodic function, i.e., for fixed $\alpha, b, \gamma, \omega$ there are multiple values of τ that satisfy (22).

4. Taking the appropriate derivatives, we find

$$\frac{d\lambda}{d\tau} = \frac{\pm\lambda c e^{-\lambda\tau}}{1 \mp \tau c e^{-\lambda\tau} - \frac{b}{(\lambda+\gamma)^2}} \,.$$

5. Putting together the results of all steps, allows us to fill in the number of eigenvalues with positive real part in each of the subregions of the c, τ plane as shown in Fig. 1.

An alternative way to use the procedure outlined above is to set the coupling coefficient (c in (12)) to zero in step 1 and follow the same procedure, varying the coupling coefficient instead of the delay. In systems with multiple delays, the procedure can be followed by setting one of the delays to zero, see e.g. (Campbell et al., 2006, 2005), for examples of this.

To close, we note the work of Olgac and Sipahi (2002, 2005) who have found a way to automate this procedure using a transformation of the characteristic equation.

2.2 Bifurcations

As noted in the previous subsection, points in parameter space where the characteristic equation has an eigenvalue with zero real part are points where the stability of an equilibrium point may change. These are places where a bifurcation may occur. As discussed elsewhere in this volume (Breakspear and Jirsa, 2006), bifurcations may lead to the creation of other equilibrium points or of a periodic orbit. We refer the reader that chapter for more background on bifurcations.

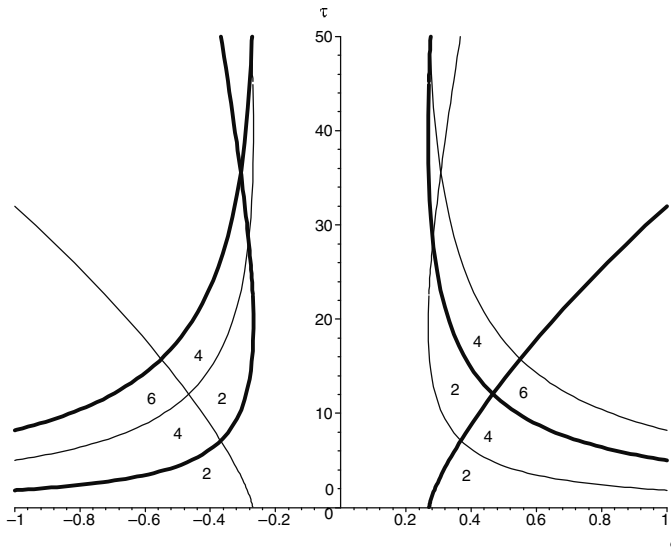


Fig. 1. Illustration of the stability and bifurcation results for the example of (12). The equilibrium solution is stable in the region contiguous with the τ axis. The number of eigenvalues with positive real part is shown in each subregion of the plane. Thick/thin curves correspond to Hopf bifurcations giving rise to synchronous/ anti-phase oscillation

Recall that the equilibrium points of (13) with $\tau > 0$ are the same as those with $\tau = 0$. Thus for the neural model (13) with $\tau > 0$, the bifurcations involving only equilibrium points (saddle-node, pitchfork, transcritical) will be the same as those for (13) with $\tau = 0$.

The two main bifurcations leading to the creation of periodic orbits in neural systems are the Hopf bifurcation and the infinite period bifurcation. These bifurcations are associated with Type II and Type I oscillators, respectively (Breakspear and Jirsa, 2006).

Consider first the Hopf bifurcation. This involves the creation of a periodic orbit as an equilibrium point changes stability. There are simple criteria to check to determine if a Hopf bifurcation occurs in a delay differential equation at a particular parameter value, say $\tau = \tau_c$.

Hopf Bifurcation Test

Assume that system (13) has an equilibrium point $\bar{\mathbf{x}}$. If the following are satisfied, then system (13) undergoes a Hopf bifurcation at $\bar{\mathbf{x}}$ as τ passes through τ_c.

1. The characteristic (19) of the linearization of (13) about $\bar{\mathbf{x}}$ has a pair of pure imaginary eigenvalues, $\pm i\omega$ when $\tau = \tau_c$, that is,

$$\Delta(\pm i\omega)\big|_{\tau=\tau_c} = 0\ .$$

2. As τ passes through τ_c the rate of change of the real part of this eigenvalue(s) is nonzero, that is, $\frac{dRe(\lambda)}{d\tau}\Big|_{\tau=\tau_c} \neq 0$.
3. The characteristic (19) of the linearization of (13) about $\bar{\mathbf{x}}$ has no other eigenvalues with zero real part.

Other than in some exceptional cases, this is enough to guarantee that a periodic orbit is created as τ passes through τ_c.

Whether the periodic orbit is stable or unstable depends on the nonlinear terms in the equation. There are two main approaches for determining this analytically, both of which require intensive computations and are best done either numerically or with a symbolic algebra package such as Maple. The centre manifold construction reduces the system of delay differential equations to a system of two ordinary differential equations from which the stability of the periodic orbit (for τ close to τ_c) may be deduced. See (Bélair et al., 1996; Wischert et al., 1994; Wu et al., 1999) for examples of how this is done. Perturbation methods, such as averaging and the method of multiple scales, find an approximate expression for the periodic solution and for the corresponding Floquet exponents. See (Campbell et al., 2006; Gopalsamy and Leung, 1996; Wirkus and Rand, 2002) for examples of how this is done.

Example. Applying this test to our coupled Fitzhugh-Nagumo model shows that the system has a Hopf bifurcation along each of the curves where the characteristic equation has a pair of pure imaginary eigenvalues, i.e., along the curves defined by (21)–(22) and shown in Fig. 1. By analyzing the solutions of the linearization (16) that correspond to the roots, one can show that some of the Hopf bifurcations give rise to synchronous or in-phase oscillations (i.e. $v_1(t) = v_2(t)$ and $w_1(t) = w_2(t)$ for all t) and some to anti-phase solutions (i.e. the spikes in v_1 and v_2 are half a period apart and similarly for w_1 and w_2).

One important thing to note about Hopf bifurcation in systems of delay differential equations is that there are always multiple branches of Hopf bifurcation. This can be seen in our example. The τ value where a Hopf bifurcation occurs corresponds to a τ value satisfying (22). Clearly if a given value of τ satisfies this equation, then so does $\tau + k\pi,\ k = \pm 1, \pm 2, \ldots$.

Now consider the the infinite period bifurcation. This bifurcation occurs when a saddle-node bifurcation occurs on an invariant circle. As indicated above, the conditions for the saddle-node bifurcation to occur in a delay differential equation are the same as for the corresponding system with zero delay. Whether or not this bifurcation occurs on a limit cycle is not easily determined analytically (even without delays), thus these bifurcations are often investigated using numerical tools (see Sect. 2.5).

2.3 Lyapunov Theory

The basic idea of Lyapunov theory is to use an auxiliary function to determine the dynamics of a nonlinear system. A very simple example is the total energy in a mechanical system with damping, such as the pendulum model:

$$\ddot{\theta} + \gamma\dot{\theta} + \frac{g}{l}\sin\theta = 0 \ .$$

The total energy of this system is

$$E(\theta, \dot{\theta}) = \frac{1}{2}\dot{\theta}^2 + gl(1 - \cos\theta) \ .$$

A simple calculation, keeping in mind that θ and $\dot{\theta}$ depend on t, shows that $\frac{dE}{dt} < 0$. This means that as t increases, E must tend to a minimum value. This in turn determines what the solutions of the nonlinear model can do. In particular, one can show that this implies that all solutions must tend to one of the equilibrium points $(\theta, \dot{\theta}) = (2k\pi, 0),\ k \in \mathbb{Z}$ as $t \to \infty$, i.e. the pendulum swings with smaller and smaller amplitude until it is hanging straight down. Lyapunov theory generalizes this idea to an arbitrary auxiliary function, $V(\mathbf{x})$, which has similar properties to the energy function in the above example, viz.,

1. $V(\mathbf{x}) > 0,\ \mathbf{x} \neq 0; V(0) = 0$ (V positive definite)
2. $\frac{dV}{dt} < 0,\ \mathbf{x} \neq 0$ ($\frac{dV}{dt}$ negative definite).

These properties can be used to show that the equilibrium point $\mathbf{x} = 0$ is asymptotically stable. By modifying the properties above, one can also use Lyapunov functions to show that an equilibrium point is unstable, that all solutions are bounded or that all solutions synchronize as $t \to \infty$.

There are two ways of extending the Lyapunov theory for ordinary differential equations to delay differential equations such as (13). **Lyapunov functionals** are auxiliary functions which depend on the value of the state over an interval in time, i.e., $V(\mathbf{x}_t)$, where $\mathbf{x}_t(\theta) = \mathbf{x}(t + \theta),\ -\tau \leq \theta \leq 0$.

The conditions for showing an equilibrium point is stable are basically the same as those outlined for the ODE case, above. The main difference comes in showing those conditions are satisfied, which can be more complicated. The **Razumikhin approach** uses an auxiliary *function* $V(\mathbf{x}(t))$, but the second condition is relaxed to $\frac{dV}{dt} < 0$ whenever $V(\mathbf{x}(t)) > V(\mathbf{x}(t + \theta)),\ -\tau \leq \theta \leq 0$. Essentially, this requires that V not increase for time intervals longer than the delay.

2.4 Phase Models

Many of the analytical tools I have discussed so far are used for studying the stability of equilibrium points and the creation of oscillatory solutions as parameters are varied. These tools are most helpful for predicting the behaviour of systems where the individual neurons do not exhibit oscillatory behaviour

when they are uncoupled. For systems which are inherently oscillatory, i.e. systems where the individual neurons exhibit oscillatory behaviour when they are uncoupled, one of the primary tools available is the phase model. The basic idea of this approach is that for a group of oscillating neurons which are weakly coupled, the key variables of importance in understanding how the neurons affect each other are the phases of the oscillators associated with the neurons. Thus a system of k model neurons, each represented by an n-dimensional system of differential equations, can be reduced to a system of k differential equations for the phases of the k oscillators. Typically these equations are in the form

$$\dot{\theta}_i(t) = \Omega t + \epsilon H_i(\mathbf{\Theta}_i(t) - \theta_i(t)\hat{\mathbf{e}})$$

where $\Theta_i(t) = (\theta_1(t), \ldots, \theta_{i-1}(t), \theta_{i+1}(t), \ldots, \theta_k(t))$, $\hat{\mathbf{e}} = (1, 1, \ldots, 1)$, Ω is the network frequency, and ϵ is the strength of the coupling. Since the coupling is weak, ϵ is small, i.e., $0 < \epsilon << 1$.

The procedure to calculate the phase model for a particular differential equation is described in Hoppensteadt and Izhikevich (1997). In most cases it is not possible to carry out this procedure analytically, however, a numerical implementation is available in the package XPPAUT (Ermentrout, 2005) and described in the book of Ermentrout (2002). The numerical implementation yields a numerical approximation of the functions H_i. A Fourier series representation of these functions can also be calculated.

There are two main results concerning phase models for equations such as (13) which have an explicit time delay in the coupling. The analysis of Ermentrout (1994) and Kopell and Ermentrout (2002) indicates that explicit time delays will produce phase shifts in the corresponding phase models provided that the delay is not a multiple of the oscillation period. Specifically, the models have the form

$$\dot{\theta}_i(t) = \Omega t + \epsilon H_i(\mathbf{\Theta}_i(t) - \theta_i(t)\hat{\mathbf{e}} - \psi) \ ,$$

where $\psi = \tau\Omega \bmod 2\pi$.

Izhikevich (1998) has refined this analysis. He has shown that Ermentrout's analysis only holds for delays as large as the order of the oscillation period, i.e., $\tau \sim 1/\Omega$. For larger delays, i.e., $\tau \sim 1/(\Omega\epsilon)$, an explicit delay will occur in the phase model. In this case the phase model will consist of a set of k delay differential equations of the form

$$\dot{\theta}_i(t) = \Omega t + \epsilon H_i(\mathbf{\Theta}_i(t - \tau) - \theta_i(t)\hat{\mathbf{e}})$$

For equations with a distributed delay in the coupling, Ermentrout (1994) and Kopell and Ermentrout (2002) have shown that the phase model will be of the form

$$\dot{\theta}_i(t) = \Omega t + \epsilon \int_0^\infty [H_i(\mathbf{\Theta}_i(t - s) - \theta_i(t)\hat{\mathbf{e}})\, g(s)]\ ds \ .$$

2.5 Numerical Tools

There are two basic numerical tools which can aid in the study of delay differential equations such as (13): numerical simulation and numerical bifurcation analysis.

In numerical simulation one attempts to determine an approximate solution of a differential equation given a particular initial state. Note that to solve such a problem for a delay differential equation such as (13), one needs to specify the value of the variable $\mathbf{x}$ not just at the start time $t = 0$, but for the whole interval $[-\tau, 0]$. Thus an initial condition for (13) is

$$\mathbf{x}(t) = \phi(t),\ -\tau \leq t \leq 0\ .$$

Typically ϕ is taken to be a constant, i.e.,

$$\mathbf{x}(t) = \mathbf{x}_0,\ -\tau \leq t \leq 0\ ,$$

which is reasonable for most experimental systems. It should be noted that only solutions which are asymptotically stable can be accurately approximated using numerical integration.

There are two main programs available for the numerical integration of delay differential equations. The widely-used (and free) package XPPAUT (Ermentrout, 2005) can perform numerical integration using a variety of fixed step numerical methods, including Runge-Kutta. It has a good graphical user interface for visualizing the results. Perhaps the most useful aspect of this program is the ease with which parameters and initial conditions can be changed. The recent book of Ermentrout (2002) gives a overview of the package including many examples. Information on how to download the package as well as documentation and tutorials are available at `www.math.pitt.edu/~bard/xpp/xpp.html`. Within MATLAB there is the function DDE23 (Shampine and Thompson, 2001) which is a variable step size numerical integration routine for delay differential equations. A tutorial is on this routine available at `www.mathworks.com/dde_tutorial`. Results maybe visualized using the extensive graphing tools of MATLAB.

Numerical bifurcation analysis consists of two parts, the approximation of a solution and the calculation of the stability of this solution. The approximation of a solution in a numerical bifurcation package is not done using numerical integration, but rather using numerical continuation. Numerical continuation uses a given solution for a particular parameter value to find a solution for a different (but close) parameter value. This is only easily implemented for equilibrium and periodic solutions. Both stable and unstable solutions can be found. Once an equilibrium solution is found to a desired accuracy, approximations for a finite set of the eigenvalues with the largest real part can be determined, which will determine the stability of the equilibrium point. The stability of periodic orbits can be numerically determined in a similar way. Numerical bifurcation packages generally track the stability of equilibrium points and periodic orbits, indicating where bifurcations occur.

There is one package available that does numerical bifurcation analysis for delay differential equations, DDE-BIFTOOL (Engelborghs et al., 2001). This package runs on MATLAB. An overview of the numerical methods used in this package and some examples applications can be found in the paper of Engelborghs et al. (2002). The user manual and information on how to download the package are available at
`www.cs.kuleuven.ac.be/cwis/research/twr/research/software/delay/`

3 Effects of delay

In this section I will outline some of the effects of delay that have been documented in the literature.

3.1 Creation of Oscillations

Time delays are commonly associated with type II oscillations, i.e. oscillations created by a Hopf bifurcation (Breakspear and Jirsa, 2006), for the following reason. There are many examples of systems that have a stable equilibrium point if the time delay is zero (or sufficiently small), but have oscillatory behaviour if the delay is large enough. In these systems, the oscillation is created via a Hopf bifurcation at a critical value of the delay. This is sometimes referred to as a **delay-induced oscillation**. One of the simplest examples of this is the following model for recurrent inhibition due to Plant (1981):

$$\begin{aligned}\dot{v}(t) &= v(t) - \frac{1}{3}v^3(t) - w(t) + c(v(t-\tau) - v_0)\\ \dot{w}(t) &= \rho(v(t) + a - bw(t))\ .\end{aligned}$$

This is a Fitzhugh-Nagumo model neuron with a delayed term which represents recurrent feedback. Plant considered parameters such that the system with no feedback has a stable equilibrium point and showed that this stability is maintained for the system with feedback and sufficiently small delay. He then showed that when $c < 0$ (i.e. the recurrent feedback is inhibitory), there is a Hopf bifurcation at a critical value of the delay, leading to oscillations.

3.2 Oscillator Death

One of the most publicized (Strogatz, 1998) effects of time delays is the fact that the presence of time delays in the coupling between oscillators can destroy the oscillations. This phenomenon, usually called **oscillator death** or **amplitude death** was first noted by Ramana Reddy et al. (1998), in their analysis of a simple model of type II oscillators with gap junctional coupling. Subsequently Ramana Reddy et al. (2000) observed this phenomenon experimentally in a system of two intrinsically oscillating circuits with the same

type of coupling. There are many papers related to delay induced oscillator death in the coupled oscillator literature, which I will not attempt to review here. Instead I will focus the discussion on results relevant to neural models.

The work of Ramana Reddy et al. (1998, 1999) shows that when two or more intrinsically oscillating elements are connected with gap junctional coupling of sufficient strength with a sufficiently large delay then the oscillations may be destroyed. Their work focused on systems where the elements were identical except for the frequency of the intrinsic oscillations and the coupling was all-to-all and symmetric (all the coupling coefficients were the same). Their model oscillator was just the normal form for the Hopf bifurcation. This behaviour has also been seen for a delayed, linearly coupled (i.e. (3) with no $\mathbf{x}_i(t)$ term) pair of van der Pol oscillators (Wirkus and Rand, 2002), and for a pair of Fitzhugh-Nagumo oscillators with delayed gap junctional coupling (Campbell and Smith, 2007). To my knowledge this has yet to be observed for other biophysical models of neural oscillators, however, it may be expected to occur for most type II oscillators. Atay (2003b) obtained results for a network of weakly nonlinear oscillators with a symmetric connection matrix and gap junctional coupling. He showed that if the intrinsic frequency of the oscillators is sufficiently similar then oscillator death can occur.

Several studies have shown that the type of oscillator death described above does not occur for type II oscillators with sigmoidal coupling (Burić and Todorović, 2003; Campbell et al., 2004; Shayer and Campbell, 2000). However, a different type of oscillator death can occur (Burić and Todorović, 2003; Burić et al., 2005; Campbell et al., 2004; Shayer and Campbell, 2000): for elements which are intrinsically excitable (i.e. not oscillating when decoupled), oscillations induced by instantaneous coupling may be lost if a time delay is introduced.

The work of Burić et al. (2005) has shown that for the type I oscillator of Terman and Wang (1995), there is no oscillator death of this latter type with either gap junctional or sigmoidal coupling. Their work also suggests that delay induced oscillator death of the first type is not possible.

The study of type II oscillator death in coupled neural systems combines various techniques of Sect. 2. Oscillator death can occur when increasing the time delay causes the *stabilization* of an equilibrium point. Values of the delay where this occurs will correspond to places where the characteristic (19) has an eigenvalue with zero real part and $\frac{dRe(\lambda)}{d\tau} < 0$. To have oscillator death, however, one must also show that the periodic orbit is eliminated. This means that at the value of τ where the equilibrium point stabilizes, there is a "reverse" Hopf bifurcation destroying the stable limit cycle. This may be checked via numerical simulations or numerical continuations (see subsection 2.5), or by showing, as outlined in subsection 2.2, that the Hopf bifurcation is subcritical. Burić et al. (2005) and Burić and Todorović (2003, 2005) have shown that for excitable Fitzhugh-Nagumo neurons, the restabilization of the equilibrium point is not always accompanied by oscillator death. In the case that the Hopf bifurcation is subcritical, the stable oscillator may persist with

the stable equilibrium point giving a region of bistability. In their model, for larger values of τ the periodic orbit is eliminated in a saddle-node bifurcation of limit cycles, leading to oscillator death.

The results of Burić et al. on type I oscillator death are primarily based on numerical simulations. To my knowledge there has been virtually no mathematical study of this situation. Recall that type I oscillators are those where the oscillation is created by an infinite period bifurcation (Breakspear and Jirsa, 2006). If such a bifurcation takes place in the coupled system with no time delay, introducing a time delay will not change the presence of the saddle-node bifurcation, however, it may affect whether this bifurcation occurs on an invariant circle. Continuity arguments would suggest that for sufficiently small delay, the saddle-node bifurcation will still occur on the invariant circle, leading to the creation of a periodic orbit at exactly the same bifurcation point as for the undelayed system. What happens for large delay remains to be investigated.

3.3 Attractor Switching and Multistability

A significant observation about ANNs of the form (14), is that many intersections between different Hopf bifurcation curves and between Hopf bifurcation curves and pitchfork bifurcation curves can occur (Bélair et al., 1996; Shayer and Campbell, 2000; Yuan and Campbell, 2004). Figure 1 shows that this occurs in our coupled Fitzhugh-Nagumo model as well. These intersection points are called **codimension two bifurcation points**. Such points can lead to more complicated dynamics including: the existence of solutions with multiple frequencies (quasiperiodicity), the coexistence of more than one stable solution (multistability) or the switching of the system from one type of solution to another as a parameter is varied (Guckenheimer and Holmes, 1983, Chap. 7), (Kuznetsov, 1995, Chap. 8). In ordinary differential equations, such points are quite rare. In delay differential equations, however, such points are more common as the time delay forces there to be multiple branches of Hopf bifurcation.

In the ANN models, the following behaviour associated with the codimension two points has been observed (Bélair et al., 1996; Campbell et al., 2005; Shayer and Campbell, 2000; Yuan and Campbell, 2004): (i) multistability between a periodic solution and one or more equilibrium points; (ii) bistability between two periodic solutions (both synchronous or one synchronous and one asynchronous); and (iii) switching from one stable solutions to another as the delay is changed for a fixed coupling strength or as the coupling strength is changed for a fixed delay. The switching in (iii) may take place through a region of bistability or a region where the trivial solution is stable. Note that situation (i) leads to a different type of oscillator death than that discussed in the previous subsection: a slight perturbation can cause the system to switch from the stable oscillatory solution to the stable equilibrium solution, with *no change* in the parameter values.

Most of this behaviour has been confirmed in systems with biophysically relevant models for the neurons. In their studies of rings of Fitzhugh-Nagumo oscillators with time delayed gap-junctional or sigmoidal coupling, Burić and Todorović (2003) and Burić et al. (2005) have documented almost all the behaviour observed in the ANN models including switching between different oscillation patterns and bistability between different oscillation patterns. For a system of two van der Pol oscillators with linear delayed coupling (i.e. (3) with no $\mathbf{x}_i(t)$ term), Sen and Rand (2003) have numerically observed and Wirkus and Rand (2002) have analytically proven the following sequence as the time delay is increased: in-phase oscillations $\rightarrow$ bistability between in-phase and anti-phase oscillations $\rightarrow$ anti-phase oscillations. They also observed the reverse sequence for different values of the coupling strength. Delay-induced bistability between in-phase oscillations and suppression oscillations (i.e. one cell oscillates and the other is quiescent) has been observed in models of hippocampal interneurons (Skinner et al., 2005a,b). Here the delay was synaptic and modelled via an extra equation representing the chemical kinetics of the synapse. Bistability between different types of travelling pulses has been observed in certain integrate-and-fire networks with delayed excitatory synaptic connections (Golomb and Ermentrout, 1999, 2000). In particular, they observe a switch from continuous travelling pulses to lurching travelling pulses as the time delay is increased with a transition region where there is bistability between the two types. This behaviour seems to be associated with a subcritical Hopf bifurcation.

Foss et al. (1996) and Milton and Foss (1997) have studied multistability in models for a delayed recurrent neural loop. Their model consists of a single excitatory neuron with delayed inhibitory feedback. They showed that up to three stable oscillatory patterns can coexist and that switching between the attractors can be induced by small perturbations in the neuron voltage (Foss et al., 1996) or by noise (Foss et al., 1997). These results have been replicated in experimental studies of a hybrid neural computer device consisting of an *Aplysia* motorneuron dynamically clamped to a computer which provides the delayed feedback (Foss and Milton, 2000, 2002). A possible cause of the multistability in these delayed feedback systems maybe period doubling bifurcations (Ikeda and Matsumoto, 1987). Bistability between different oscillation patterns was also observed in preparations of small *Aplysia* neural circuits (Kleinfeld et al., 1990).

Bifurcation induced transitions between different attractors have been observed in several experiments. In an experimental electrical circuit system, Ramana Reddy et al. (2000) have observed the sequence: in-phase oscillations $\rightarrow$ no oscillations $\rightarrow$ anti-phase oscillations as the time delay in the (gap-junctional) coupling is increased. Transitions from in-phase to anti-phase oscillations have been observed in human bimanual coordination experiments (Kelso et al., 1981; Kelso, 1984; Carson et al., 1994); see also the review article of Jantzen and Kelso (2006). One model which explains these experiments incorporates time delays in the coupling (Haken et al., 1985).

3.4 Synchronization

There are several approaches to studying synchronization. I will not review the details here, but give some indication which of these have been extended to delay differential equations and what the results are.

There is a very large literature on synchronization in artificial neural networks, some of which addresses systems with time delays (Campbell et al., 2006; Wu et al., 1999; Yuan and Campbell, 2004; Zhou et al., 2004a,b). Most of these papers use Lyapunov functionals to show that the all solutions synchronize as $t \to \infty$, for appropriate parameter values. Although the equations of the individual elements are not relevant for modelling biophysical neurons, the techniques of analysis may be carried over to neural systems. A common conclusion in many of these papers is that if the strength of the coupling is small enough, one can achieve synchronization for all $\tau \geq 0$. However, synchronization may mean that all elements asymptotically approach the same equilibrium point.

As I have mentioned elsewhere in this chapter, a basic principle of delay differential equations such as (13) is that the behaviour of the system for small delay is often qualitatively similar to that for zero delay. Thus if the neurons are synchronized for a given value of the coupling with zero delay they should remain synchronized for small enough delays in the coupling. Unfortunately, quantifying "small enough" may be difficult and will generally depend on the particular neural model involved. Recall the example illustrated in Fig. 1. We showed that for $c > 0$ large enough (sufficiently large excitatory coupling) the undelayed system exhibits synchronized oscillations. We expect these oscillations to persist for $\tau > 0$ at least until one reaches the first thick Hopf bifurcation curve where synchronous oscillations are destroyed. (If the Hopf bifurcation is subcritical, the oscillations may persist above the curve). Thus, for this particular example, the Hopf bifurcation curve gives a lower bound on "how small" the delay must be to preserve the synchronization found for zero delay. Note that this does not preclude synchronization occurring for larger values of the delay, which is the case in this example. A similar situation is seen for coupled van der Pol oscillators in (Wirkus and Rand, 2002). Another example is the work of Fox et al. (2001) who studied relaxation oscillators with excitatory time delayed coupling. They showed that synchrony achieved for zero delay is preserved for delays up to about 10% of the period of the oscillation, for a variety of different models. The one exception is when the right hand side of the equation is not a differentiable function, in which case synchronization is lost for $\tau > 0$. Crook et al. (1997) observed a similar phenomenon for a continuum model of the cortex, with excitatory coupling and distance dependent delays. Namely, they found for small enough delay the synchronous oscillation is stable, but for larger delays this oscillation loses stability to a travelling wave.

More complicated situations occur when both excitatory and inhibitory connections exist. Ermentrout and Kopell (1998); Kopell et al. (2000); Karbowski and Kopell (2000) have studied a model for hippocampal networks of excitatory and inhibitory neurons where two types of synchronous oscillation are possible. They show that persistence of the synchronous oscillations with delays depends subtly on the currents present in the cells and the connections present between cells.

So far I have discussed synchronization *in spite of* delays. I now move on to the more interesting case of synchronization *because of* delays. This situation can occur when there are inhibitory synaptic connections in the network. This has been extensively documented and studied when the delay is modelled by slow kinetics of the synaptic gating variable (van Vreeswijk et al., 1994; Wang and Buzsáki, 1998; Wang and Rinzel, 1992, 1993; White et al., 1998). Further, Maex and De Schutter (2003) suggest that the type of delay is not important, just the fact that it leads to a separation in time between when the pre-synaptic neuron generates an action potential and the post-synaptic neuron receives it. They confirm this for a network of multi-compartment model neurons with fast synaptic kinetics and a discrete conduction delay. This idea is further supported by the observation of synchronization via discrete delayed inhibition in a number of artificial neural network models (Campbell et al., 2004, 2005; Shayer and Campbell, 2000). Finally we illustrate this with our coupled Fitzhugh-Nagumo model. Consider the part of Fig. 1 with $c < 0$ (inhibitory coupling). For sufficiently large coupling strength and zero delay the system tends to an asynchronous phase-locked state. This state persists for $\tau > 0$ sufficiently small, however, for τ large enough a stable synchronous state may be created in the Hopf bifurcation corresponding to the thin curve.

Only a few studies have looked at synchronization with time delayed gap-junctional coupling. One example is the work of Dhamala et al. (2004) which shows that for two gap junctional coupled Hindmarsh-Rose neurons synchronization is achieved for smaller coupling strengths if there is a nonzero time delay in the coupling. Another is the work of Burić et al. (2005).

4 Distributed Delays

There are very few results concerning neural systems with distributed delays, thus I will review some general results, mostly from the population biology literature, which should carry over to neural systems. What has emerged from this literature is a general principle that *a system with a distribution of delays is inherently more stable than the same system with a discrete delay*. Some specific results to support this are described below.

Bernard et al. (2001) analyzed the linear stability of a scalar system with one and two delays in terms of generic properties of the distribution g, such as the mean, variance and skewness. For the uniform and continuous

distributions, they have shown that stability regions are larger than those with a discrete delay.

Jirsa and Ding (2004) have analyzed an $n \times n$ linear system with linear decay and arbitrary connections with a common delay. They have shown, under some mild assumptions, that the stability region of the trivial solution for any distribution of delays is larger than and contains the stability region for a discrete delay.

Campbell and Ncube (2006) have shown that it is more difficult to get delay induced oscillations with distributions of delays of the form (6) with $\tau_m = 0$. For large variance ($m = 1$) delay induced instability is impossible and for smaller variance ($m > 1$) the mean delay needed for instability is much larger than the discrete delay value. They have also shown that sufficiently small variance in the distribution is needed to get the bifurcation interactions which may lead to multistability, oscillator death and attractor switching discussed above.

Atay (2003a, 2006) has studied the same model as Ramana Reddy et al. (1998) only with distributed delays of the form (6) with g given by (7). He shows it is easier to destroy oscillations with a distribution of delays than with a discrete delay, in the sense that there is a larger region of oscillator death in the parameter space consisting of the mean delay and the strength of the coupling. As the variance of the distribution increases the size of this region increases.

Thiel et al. (2003) studied a scalar equation representing a mean field approximation for a population pyramidal cells with recurrent feedback, first formulated by Mackey and an der Heiden (1984). They show that having a uniform distribution of delays simplifies the dynamics of the system. The size of the stability region of the equilibrium point is larger and larger mean delays are needed to induce oscillations. Complex phenomena such as chaos are less likely to occur, or totally precluded if the variance of the distribution is sufficiently large. The model with a distribution of delays better explains the appearance of periodic bursts of activity when penicillin is added to a hippocampal slice preparation (which reduces the coupling strength).

5 Summary and Future Directions

In this chapter I showed how time delays due to conduction along the axon or dendrite or due to transmission across the synapse could be modelled with delay differential equations. I outlined some of the tools available for analyzing such equations and reviewed some of the literature about such models. Some key observations are:

- Time delays can lead to the creation of type II oscillations, especially in systems with delayed inhibitory coupling.
- Time delays can destroy type II oscillations in a network of intrinsically oscillatory neurons with gap junctional coupling.

- If a system has a stable synchronous oscillation when there is no delay in the coupling, the solution remains stable for small enough delay, but may lose stability for larger delay.
- A system with inhibitory coupling which does not have a stable synchronous oscillation for zero delay, may have one if the delay is large enough.
- Time delays may lead to bistability between different type II oscillatory solutions (e.g. synchronous and anti-phase) or switching between different type II oscillatory solutions.

There are a number of problems which still require further study. These include: determining the effect of delay on the generation and destruction of type I oscillations (infinite period bifurcations), applying and/or extending the methods used to study synchronization in artificial neural networks to biophysical neural networks, and studying the effect of distributions of delays on biophysical neural networks.

References

Atay FM (2003a) Distributed delays facilitate amplitude death of coupled oscillators. Phys. Rev. Lett. 91 (9), 094101

Atay FM (2003b) Total and partial amplitude death in networks of diffusively coupled oscillators. Physica D 183, 1–18

Atay FM (2006) Oscillator death in coupled functional differential equations near Hopf bifurcation. J. Differential Equations 221 (1), 190–209

Bélair J, Campbell SA, van den Driessche P (1996) Frustration, stability and delay-induced oscillations in a neural network model. SIAM J. Applied Mathematics 56 (1), 245–255

Bernard S, Bélair J, Mackey MC (2001) Sufficient conditions for stability of linear differential equations with distributed delay. Discrete and Continuous Dynamical Systems 1B, 233–256

Breakspear M, Jirsa VK (2006) Neuronal dynamics and brain connectivity. In: McIntosh R, Jirsa VK (eds), Handbook of Brain Connectivity. Springer-Verlag, New York

Burić N, Grozdanović I, Vasović N (2005) Type I vs. type II excitable systems with delayed coupling. Chaos, Solitons and Fractals 23, 1221–1233

Burić N, Todorović D (2003) Dynamics of Fitzhugh-Nagumo excitable systems with delayed coupling. Physical Review E 67, 066222

Burić N, Todorović D (2005) Bifurcations due to small time-lag in coupled excitable systems. International Journal of Bifurcations and Chaos 15 (5), 1775–1785

Campbell SA, Edwards R, van den Dreissche P (2004) Delayed coupling between two neural network loops. SIAM J. Applied Mathematics 65 (1), 316–335

Campbell SA, Ncube I (2006) Some effects of gamma distribution on the dynamics of a scalar delay differential equation. Preprint

Campbell SA, Ncube I, Wu J (2006) Multistability and stable asynchronous periodic oscillations in a multiple-delayed neural system. Physica D 214 (2), 101–119

Campbell SA, Smith A (2007) Phase models and delayed coupled Fitzhugh-Nagumo oscillators. Preprint

Campbell SA, Yuan Y, Bungay S (2005) Equivariant Hopf bifurcation in a ring of identical cells with delayed coupling. Nonlinearity 18, 2827–2846

Carson RG, Byblow WD, Goodman D (1994) The dynamical substructure of bimanual coordination. In: Swinnen S, Heuer H, Massion J, Casaer P (eds), Interlimb coordination: Neural, dynamical and cognitive constraints. Academic Press, San Diego, pp 319–337

Crook SM, Ermentrout GB, Vanier MC, Bower JM (1997) The role of axonal delay in the synchronization of networks of coupled cortical oscillators. J. Computational Neuroscience 4, 161–172

Cushing JM (1977) Integrodifferential Equations and Delay Models in Popluation Dynamics. Vol. 20 of Lecture Notes in Biomathematics. Springer-Verlag, Berlin; New York

Desmedt JE, Cheron G (1980) Central somatosensory conduction in man: Neural generators and interpeak latencies of the far-field components recorded from neck and right or left scalp and earlobes. Electro. Clin. Electroencephalog. 50, 382–403

Dhamala M, Jirsa VK, Ding M (2004) Enhancement of neural synchrony by time delay. Physical Review E 92 (7), 074104

Diekmann O, van Gils SA, Verduyn Lunel SM, Walther H-O (1995) Delay Equations. Springer-Verlag, New York

Engelborghs K, Luzyanina T, Roose D (2002) Numerical bifurcation analysis of delay differential equations using DDE-BIFTOOL. ACM Transactions on Mathematical Software 28 (1), 1–21

Engelborghs K, Luzyanina T, Samaey G (2001) DDE-BIFTOOL v. 2.00: a MATLABpackage for bifurcation analysis of delay differential equations. Tech. Rep. TW-330, Department of Computer Science, K.U. Leuven, Leuven, Belgium

Ermentrout GB (1994) An introduction to neural oscillators. In: Ventriglia F (ed), Neural Modeling and Neural Networks. Pergamon, Oxford, UK, pp 79–110

Ermentrout GB (2002) Similating, Analyzing and Animating Dynamical Systems: A Guide to XPPAUT for Researcher and Students. SIAM, Philadelphia, PA

Ermentrout GB (2005) XPPAUT5.91 – the differential equations tool. Department of Mathematics, University of Pittsburgh, Pittsburgh, PA

Ermentrout GB, Kopell N (1998) Fine structure of neural spiking and synchronization in the presence of conduction delays. PNAS 95 (3), 1259–64

Fitzhugh R (1960) Thresholds and plateaus in the Hodgkin-Huxley nerve equations. J. General Physiology 43, 867

Foss J, Longtin A, Mensour B, Milton JG (1996) Multistability and delayed recurrent loops. Phys. Rev. Letters 76, 708–711

Foss J, Milton JG (2000) Multistability in recurrent loops arising from delay. J. Neurophysiol. 84, 975–985

Foss J, Milton JG (2002) Multistability in delayed recurrent loops. In: Milton J, Jung P (eds), Epilepsy as a Dynamic Disease. Springer-Verlag, New York, pp 283–295

Foss J, Moss F, Milton JG (1997) Noise, multistability and delayed recurrent loops. Phys. Rev. E 55, 4536–4543

Fox JJ, Jayaprakash C, Wang DL, Campbell SR (2001) Synchronization in relaxation oscillator networks with conduction delays. Neural Computation 13, 1003–1021

Golomb D, Ermentrout GB (1999) Continuous and lurching travelling pulses in neuronal networks with delay and spatially decaying connectivity. PNAS 96, 13480–13485

Golomb D, Ermentrout GB (2000) Effects of delay on the type and velocity of travelling pulses in neuronal networks with spatially decaying connectivity. Network: Comput. Neural Syst. 11, 221–246

Gopalsamy K, Leung I (1996) Delay induced periodicity in a neural netlet of excitation and inhibition. Physica D 89, 395–426

Guckenheimer J, Holmes PJ (1983) Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields. Springer-Verlag, New York

Haken H, Kelso JAS, Bunz H (1985) A theoretical model of phase transitions in human hand movements. Biological Cybernetics 51, 347–356

Hale JK, Verduyn Lunel SM (1993) Introduction to Functional Differential Equations. Springer-Verlag, New York

Hoppensteadt FC, Izhikevich EM (1997) Weakly connected neural networks. Springer-Verlag, New York

Ikeda I, Matsumoto K (1987) High dimensional chaotic behaviour in systems with time-delayed feedback. Physica D 29, 223–235

Izhikevich EM (1998) Phase models with explicit time delays. Physical Review E 58 (1), 905–908

Jantzen KJ, Kelso JAS (2006) Neural coordination dynamics of human sensorimotor behaviour: A review. In: McIntosh R, Jirsa VK (eds), Handbook of Brain Connectivity. Springer-Verlag, New York

Jirsa VK, Ding M (2004) Will a large complex system with delays be stable? Physical Review Letters 93, 070602

Karbowski J, Kopell N (2000) Multispikes and synchronization in a large neural network with temporal delays. Neural Computation 12, 1573–1606

Keener J, Sneyd J (1998) Mathematical Physiology. Springer-Velag, New York

Kelso JAS (1984) Phase transitions and critical behaviour in human bimanual coordination. American J. Physiology: Regulatory, Integrative and Comparative Physiology 15, R1000–R1004

Kelso JAS, Holt KG, Rubin P, Kugler PN (1981) Patterns of human interlimb coordination emerge from nonlinear limit cycle oscillatory processes: theory and data. J. Motor Behaviour 13, 226–261

Kleinfeld D, Raccuia-Behling F, Chiel HJ (1990) Circuits constructed from identified Aplysia neurons exhibit multiple patterns of activity. Biophysical J. 57 (4), 697–715

Koch C (1999) Biophysics of Computation. Oxford University Press, New York

Kolmanovskii VB, Nosov VR (1986) Stability of functional differential equations. Vol. 180 of Mathematics in Science and Engineering. Academic Press

Kopell N, Ermentrout GB (2002) Mechanisms of phase-locking and frequency control in pairs of coupled neural oscillators. In: Fiedler B (ed), Handbook of Dynamical Systems, vol 2: Toward Applications. Elsevier, Amsterdam

Kopell N, Ermentrout GB, Whittington MA, Traub R (2000) Gamma rhythms and beta rhythms have different synchronization properties. PNAS 97 (4), 1867–1872

Kuznetsov YA (1995) Elements of Applied Bifurcation Theory. Vol. 112 of Applied Mathematical Sciences. Springer-Verlag, Berlin; New York

MacDonald N (1978) Time lags in biological models. Vol. 27 of Lecture notes in biomathematics. Springer-Verlag, Berlin; New York

Mackey MC, an der Heiden U (1984) The dynamics of recurrent inhibition. J. Mathematical Biology 19, 211–225

Maex R, De Schutter E (2003) Resonant synchronization in heterogeneous networks of inhibitory neurons. J. Neuroscience 23 (33), 10503–10514

Milton JG, Foss J (1997) Oscillations and multistability in delayed feedback control. In: Othmer HG, Adler FR, Lewis MA, Dallon JC (eds), The Art of Mathematical Modeling: Case Studies in Ecology, Physiology and Cell Biology. Prentice Hall, New York, pp 179–198

Nagumo J, Arimoto S, Yoshizawa S (1962) An active pulse transmission line simulating nerve axon. Proc. IRE 50, 2061–2070

Olgac N, Sipahi R (2002) An exact method for the stability analysis of time-delayed linear time-invariant (LTI) systems. IEEE Transactions on Automatic Control 47 (5), 793–797

Olgac N, Sipahi R (2005) Complete stability robustness of third-order LTI multiple time-delay systems. Automatica 41, 1413–1422

Plant RE (1981) A Fitzhugh differential-difference equation modeling recurrent neural feedback. SIAM J. Applied Mathematics 40 (1), 150–162

Ramana Reddy DV, Sen A, Johnston GL (1998) Time delay induced death in coupled limit cycle oscillators. Physical Review Letters 80, 5109–5112

Ramana Reddy DV, Sen A, Johnston GL (1999) Time delay effects on coupled limit cycle oscillators at Hopf bifurcation. Physica D 129, 15–34

Ramana Reddy DV, Sen A, Johnston GL (2000) Experimental evidence of time-delay-induced death in coupled limit-cycle oscillators. Physical Review Letters 85 (16), 3381–3384

Sen AK, Rand R (2003) A numerical investigation of the dynamics of a system of two time-delayed coupled relaxation oscillators. Communications on Pure and Applied Mathematics 2 (4), 567–577

Shampine LF, Thompson S (2001) Solving DDEs in MATLAB. Applied Numerical Mathematics 37, 441–458

Shayer LP, Campbell SA (2000) Stability, bifurcation and multistability in a system of two coupled neurons with multiple time delays. SIAM J. Applied Mathematics 61 (2), 673–700

Shepherd G (1994) Neurobiology. Oxford University Press, New York

Skinner FK, Bazzazi H, Campbell SA (2005a) Two-cell to n-cell heterogeneous, inhibitory networks: precise linking of multistable and coherent properties. J. Computational Neuroscience 18 (3), 343–352

Skinner FK, Chung JYJ, Ncube I, Murray PA, Campbell SA (2005b) Using heterogeneity to predict inhibitory model characteristics. J. Neurophysiology 93, 1898–1907

Stépán G (1989) Retarded Dynamical Systems. Vol. 210 of Pitman Research Notes in Mathematics. Longman Group, Essex

Strogatz SH (1998) Death by delay. Nature 394, 316–317

Terman D, Wang DL (1995) Global competition and local cooperation in a network of neural oscillators. Physica D 81, 148–176

Thiel A, Schwegler H, Eurich CW (2003) Complex dynamics is abolished in delayed recurrent systems with distributed feedback times. Complexity 8 (4), 102–108

van Vreeswijk C, Abbott LF, Ermentrout GB (1994) When inhibition not excitation synchronizes neural firing. J. Computational Neuroscience 1, 313–321

Wang X-J, Buzsáki G (1998) Gamma oscillation by synaptic inhibition in a hippocampal interneuronal network model. J. Neuroscience 16 (20), 6–16

Wang X-J, Rinzel J (1992) Alternating and synchronous rhythms in reciprocally inhibitory model neurons. Neural Computation 4, 84–97

Wang X-J, Rinzel J (1993) Spindle rhythmicity in the reticularis thalami nucleus: synchronization among mutually inhibitory neurons. Neuroscience 53, 899–904

White JA, Chow CC, Ritt J, Soto-Treviño C, Kopell N (1998) Synchronization and oscillatory dynamics in heterogeneous, mutually inhibited neurons. J. Computational Neuroscience 5, 5–16

Wirkus S, Rand R (2002) The dynamics of two coupled van der Pol oscillators with delay coupling. Nonlinear Dynamics 30, 205–221

Wischert W, Wunderlin A, Pelster A, Olivier M, Groslambert J (1994) Delay-induced instabilities in nonlinear feedback systems. Physical Review E 49 (1), 203–219

Wu J, Faria T, Huang YS (1999) Synchronization and stable phase-locking in a network of neurons with memory. Math. Comp. Modelling 30 (1-2), 117–138

Yuan Y, Campbell SA (2004) Stability and synchronization of a ring of identical cells with delayed coupling. J. Dynamics and Differential Equations 16 (1), 709–744

Zhou J, Chen T, Lan Z (2004a) Robust synchronization of coupled delayed recurrent neural networks. In: Advances in Neural Networks - ISNN 2004. Vol. 3173 of Lecture Notes in Computer Science. Springer-Verlag, Berlin, pp 144–149

Zhou S, Liao Z, Yu J, Wong K (2004b) Chaos and its synchronization in two-neuron systems with discrete delays. Chaos, Solitons and Fractals 21, 133–142

Connectivity and Dynamics in Local Cortical Networks

John M Beggs, Jeffrey Klukas and Wei Chen

Department of Physics, Indiana University
727 East Third St., Bloomington, IN 47405-7105
jmbeggs@indiana.edu

Recent experimental work has begun to characterize activity in local cortical networks containing thousands of neurons. There has also been an explosion of work on connectivity in networks of all types. It would seem natural then to explore the influence of connectivity on dynamics at the local network level. In this chapter, we will give an overview of this emerging area. After a brief introduction, we will first review early neural network models and show how they suggested attractor dynamics of spatial activity patterns, based on recurrent connectivity. Second, we will review physiological reports of repeating spatial activity patterns that have been influenced by this initial concept of attractors. Third, we will introduce tools from dynamical systems theory that will allow us to precisely quantify neural network dynamics. Fourth, we will apply these tools to simple network models where connectivity can be tuned. We will conclude with a summary and a discussion of future prospects.

1 Introduction

The advent of fMRI and other imaging technology has spawned a deluge of research examining how the brain functions at the macroscopic level. This work, which treats each voxel as the basic unit of analysis, has yielded tremendous insights as to how networks of cortical regions cooperate to produce motor activity (Jantzen KJ et al., 2005; Rowe J et al., 2002), memory (Fletcher P et al., 1999), cognition (Mechelli A et al., 2004; Stephan KE et al., 2003) and emotion (Canli T et al., 2002). But within each voxel there lie perhaps tens of thousands of neurons that are connected into local networks, performing elementary computations that are fundamental to the brain's higher functions. Relatively little experimental work has been done at this mesoscopic level, despite the existence of a large literature on neural network theory and models. This chapter will focus on the relationship between network connectivity and dynamics at this level, with the hope that the principles uncovered here will be generally applicable to networks at larger scales as well. In addition,

this chapter will emphasize local networks of cortical neurons, since this is the area within the mesoscopic level where the most experimental work has been done.

2 Attractors in Early Models of Local Neural Networks

The simplest neural network models only have projections from one layer of neurons to the next, having what is called "feed-forward" architecture. While these models can do many impressive things, they can not exhibit dynamics in the true sense because their outputs are never fed back as inputs to the network. Their activity changes from layer to layer, but their final output is given only at one point in time. By contrast, in recurrent networks projections from some or all of the neurons are connected back to the inputs of the network through recurrent collaterals. Recurrent networks can therefore generate activity in a given layer that changes over time with each loop of processing, thus demonstrating dynamics. Since real brains are filled with recurrent and not purely feed-forward connections, it seems that recurrent networks are also much more realistic models of connectivity in living neural networks. For these reasons, we will only consider recurrent networks in what follows.

Much of the early work in recurrent network models was concerned with memory storage and retrieval. These simplified models demonstrated how groups of neurons could collectively store a spatial activity pattern embedded in connection strengths. An example of such a model is shown in Fig.1. The five pyramidal, excitatory neurons have all-to-all recurrent connections. When three of the neurons are activated at the same time, synaptic connections between the active collaterals and active neurons are strengthened. This rule for changing synaptic strengths is called the "Hebb rule" after Donald Hebb who most famously proposed it (Hebb DO, 1949), and is often summarized by the phrase "cells that fire together, wire together." Once these synaptic connections are strengthened by a Hebbian rule, the network has a distributed memory trace of the original configuration of the three active cells. The idea of synaptic strengths encoding memory is not new and can be traced back to Cajal (Ramón Y Cajal S, 1909), but the dynamics of this simple model was not appreciated until decades later. When a fragment of the original, stored configuration of cells is presented to the network, the network will have a tendency to use the fragment to reconstruct the original stored configuration. Active cells will recruit other cells from the stored pattern through recurrent collaterals and recently strengthened synaptic connections. The configuration of the network at each time step will thus become progressively more similar to the originally stored configuration. One way of describing this is to say that the network is attracted to the stored configuration. If the network configurations could be symbolized by binary strings and arrows could represent transitions

over time, we would have [00100] → [10100] → [10110]. But note that several other initial configurations could also lead to this final stored configuration. For example: [10000] → [10010] → [10110], and [00010] → [00110] → [10110] are also pathways. All of those configurations that eventually lead to the stored configuration are said to be in the basin of attraction of the stored configuration. The stored configuration [10110] is called an attractor in this network. In larger models, it is possible to have many independent configurations stored as attractors within the same network.

The model in Fig.1 is representative of a whole class of influential models that employed recurrent connectivity and Hebbian learning to store spatial patterns. A precursor of this class was proposed by Steinbuch

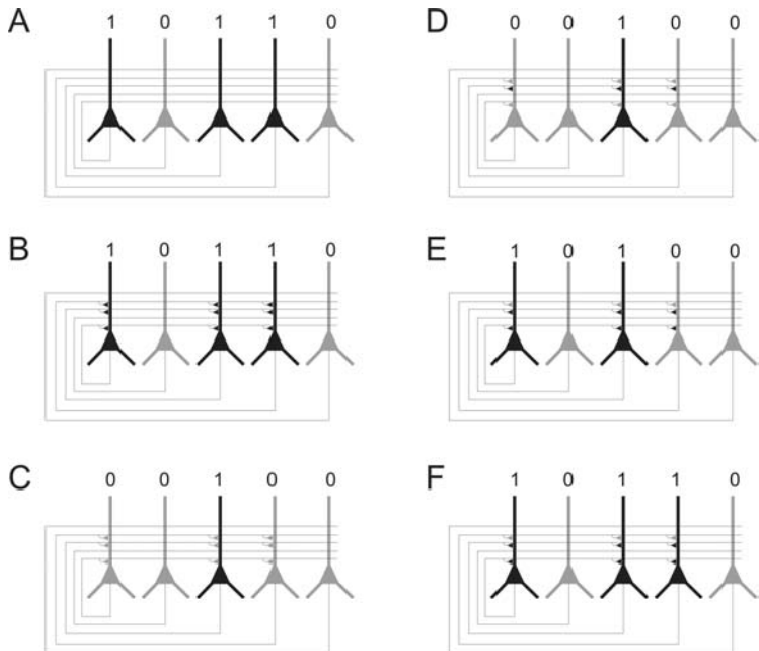


Fig. 1. An attractor in a simplified recurrent network model. Network has five pyramidal cells. Straight lines represent axon collaterals, here wired to have all-to-all connectivity. **A**, Tabula rasa: a stimulus pattern activates three neurons, shown in black. **B**, Learning: Hebbian plasticity strengthens connections between active neurons and active axon collaterals, shown as triangular synaptic connections. **C**, Cue: some time later, a fragment of the original stimulus activates the middle neuron. **D**: Beginning recall: the active neuron now drives the newly strengthened synapses, shown in black. **E**, Further recall: activity in these new synapses activates another neuron from the stimulus pattern. **F**, Total recall: collective activity now drives the third neuron from the original pattern. Over time, the state of the network became more similar to the activity pattern seen in A. Note that any partial cue of the original pattern could lead to re-activation of the original pattern. After learning, the network configuration is said to be attracted to the state shown in A

(Steinbuch K, 1961; Steinbuch K and H Frank, 1961) in his matrix memory model, which used co-activation to imprint connections and to associatively store information. Anderson's autoassociator model (Anderson JA et al., 1977), the Hopfield model (Hopfield JJ, 1982, 1984; Hopfield JJ and DW Tank, 1986) and models analyzed by Cohen and Grossberg (Cohen MA and S Grossberg, 1983) all used Hebbian learning and had all-to-all connectivity. An emergent property of these models, stemming in part from their connectivity, was that information could be stored in attractors, and that network activity would tend to settle into these attractors (Amit DJ, 1989). The models of Hopfield and Grossberg are also noteworthy for other reasons (connecting statistical physics to neural network theory; using time as an important variable in network dynamics) that are beyond the scope of this chapter. For our purposes, it is important to note that these models used recurrent connections and proposed that spatial information could be stored in attractors. Versions of this class of model were later elaborated by neuroscientists to explain how the hippocampus might store and retrieve memories (Rolls ET, 1990; Skaggs WE and BL McNaughton, 1992).

It is also worth noting that much work has been done on how even single neurons with recurrent connections can store *temporal* information in spike sequences (e.g., Foss, Longtin, Mensour and Milton, 1996; Foss and Milton, 2000). These sequences can be considered attractors, although they may not necessarily store spatial patterns of activity across many neurons, as we have been discussing. For further coverage of this interesting topic, the reader is referred to Sue Ann Campbell's chapter in this handbook.

The simple class of models which store spatial patterns of activity was appealing to computational neuroscientists for several reasons. First, it seemed biologically plausible. As stated before, recurrent collaterals are abundant in the brain, and there is ample evidence that synapses can be strengthened according to a Hebbian rule (Kelso SR et al., 1986; Kirkwood A and MF Bear, 1994). Second, the dynamics of the model seem to mimic the way memories are subjectively recalled. Presenting a cue or fragment of information is often enough to elicit more detailed information that was originally associated with it. Just as viewing a fragment of a picture can often evoke a complete image from memory, so also a few active neurons can cause the model to complete the pattern that was originally stored (Hopfield JJ, 1982). Third, these models allowed several patterns to be stored within the same network, a property that clearly would be useful in real brains. Because of their plausibility and impressive emergent properties, these simple network models caused many researchers to expect that local circuits in mammalian cortex would store memories in the form of attractors.

3 Repeating Activity Patterns and the Influence of Attractor Models

Is there any evidence for attractors, as described by the above model, in physiological recordings? In order to evaluate this form of the attractor hypothesis correctly, several experimental requirements need to be met. First, since activity is hypothesized to be distributed among many neurons, multiple recording sites are needed. Second, network activity must visit some configurations more often than would be expected by chance. If all network activity configurations were visited equally often, there would be no attractors. But if some configurations are visited repeatedly and more often than would be expected by chance, then there is at least a possibility that attractors exist in the network. Third, when the network is in a configuration that is close to one of its stored configurations, network activity should become progressively more similar to the stored configuration over time. This indicates that the network is being drawn into the attractor. Fourth, these repeatable configurations need to be stable over time if they are to serve as a substrate for information storage.

Among the first recordings that fulfilled some of these requirements were those from Abeles and colleagues, who observed temporally precise spike sequences in primate cortex (Abeles M et al., 1993; Ben-Shaul Y et al., 2004). They reported that spike triplets, which they later called "synfire chains," reproducibly appeared while monkeys were engaging in particular stages of cognitive tasks. In addition, these sequences occurred more often than would be expected by chance, under the assumption that spike trains can be modeled as a random Poisson process. Although some researchers later questioned whether the synfire chains reported by Abeles and colleagues were truly statistically significant (Baker SN and RN Lemon , 2000; Oram MW et al., 1999), other groups gradually began to report repeating activity patterns as well. Recordings from rat hippocampus showed that distributed patterns of neurons became active as rats made their way through a maze (Wilson MA and BL McNaughton, 1993). Whenever a rat revisited a portion of the maze in the same way, a similar pattern of activity would appear (Brown EN et al., 1998; Skaggs WE et al., 1996). These similarities were statistically significant, and suggested that the activity configuration somehow represented spatial or cue information. Interestingly, these patterns were later found to significantly reappear during subsequent, but not previous, sleep sessions (Lee AK and MA Wilson, 2002; Louie K and MA Wilson, 2001; Nadasdy Z et al., 1999). This suggested that the activity patterns encoded the previous day's maze running session and were being consolidated during sleep (Wilson MA, 2002), a hypothesis that is still somewhat disputed. Less controversially, these data indicated that the reproducible activity patterns had long-term stability and could serve as a substrate for information storage (Lee AK and MA Wilson, 2004). Reproducible activity patterns were also found in the cortex-like structure HVC (high vocal center) of song birds during

song learning and production (Hahnloser RH et al., 2002). The temporal precision of these activity patterns like this was astoundingly high, being 1 millisecond or less (Chi Z and D Margoliash, 2001). Activity patterns observed in song birds also had long-term stability and replayed during sleep (Dave AS and D Margoliash, 2000; Deregnaucourt S et al., 2005), indicating that they too could serve to store information. Reproducible activity patterns have now been found in a variety of in vivo systems ranging from visual cortex (Kenet T et al., 2003), and the olfactory bulb (Spors H and A Grinvald, 2002) to the brain stem (Lindsey BG et al., 1997). Collectively, these data demonstrate that distributed, reproducible activity patterns with long-term stability exist in the intact brain.

But did these patterns arise because many different brain areas were acting together? It remained to be seen whether isolated portions of brain could sustain reproducible activity patterns. Yuste and colleagues used calcium dyes and a scanning two-photon microscope to image activity from hundreds of sites in acute slices of mouse visual cortex (Cossart R et al., 2003; Mao BQ et al., 2001). They reported that neurons became active in particular patterns that reoccurred more often than would be expected by chance. Because the microscope had to scan over so many neurons, it took about one second before the scanning laser could return to a given neuron to image it again. Thus, they were able to image activity over the cortical slice network at a temporal resolution of about 1 second. This exciting work demonstrated that neocortical tissue in isolation spontaneously produced repeatable activity patterns, and raised the possibility that local circuit connectivity, to the extent that it was preserved in the slice, was sufficient to support these patterns. Further evidence that local networks were enough to generate attractor-like patterns came from work with neural cultures grown on 60-channel multielectrode arrays. Using cultured slices prepared from rat cortex, Beggs and Plenz (Beggs JM and D Plenz, 2004) showed that reproducible activity patterns had a temporal precision of 4 milliseconds and were stable for as long as 10 hours (Fig. 2). While these cultures were prepared from slices that preserved some of the intrinsic cortical circuitry, they were grown for three weeks in isolation from sensory inputs. Thus, the activity patterns that arose were very likely to have been the result of self-organizing mechanisms (e.g., Hebbian rules, homeostatic regulation of firing rate) present at the neuronal and synaptic levels. As even further evidence that repeating activity patterns can result from self-organization, Ben-Jacob and colleagues (Segev R et al., 2004) have demonstrated that networks of dissociated cultures produce repeating activity patterns. These cultures are prepared from suspensions of individual neurons that are then poured over an electrode array and grown in an incubator for several weeks. As a result, these preparations do not preserve intrinsic cortical circuitry at all, even though they may match the proportions of excitatory and inhibitory cells found in cortex. Collectively, this work indicates that long-lasting, temporally precise, reproducible activity patterns can readily form in isolated cortical tissue. The fact that even dissociated cultures can generate

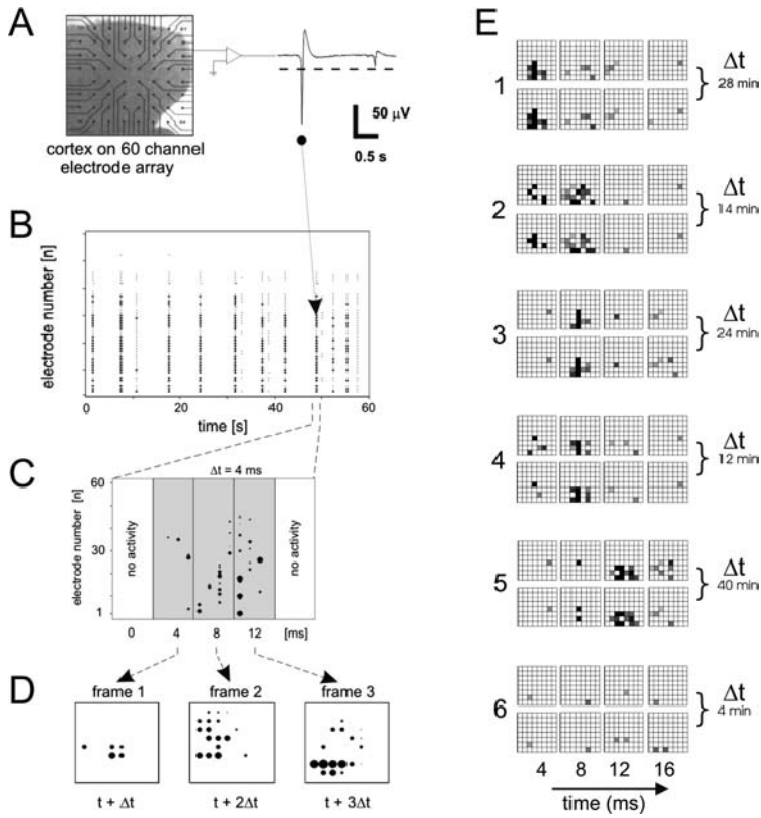


Fig. 2. Reproducible activity patterns from an isolated cortical network. **A**, On the left is an organotypic culture from rat somatosensory cortex (containing ∼50,000 neurons) pictured on a 60-channel multielectrode array at 2 days in vitro. Electrodes are seen as small black dots at the end of straight lines. Electrode tips are 30 µm in diameter and the inter-electrode distance is 200 µm. **B**, On the right is the local field potential signal recorded from one electrode, low-pass filtered at 50 Hz. The dashed line is a threshold set at −3 standard deviations. The sizes of the dots represent the magnitudes of the suprathreshold field potentials. B, The raster plot of activity from all electrodes is shown for one minute. Columns of dots indicate nearly synchronous bursts activity on many electrodes. Activity bursts are separated by quiescent intervals of several seconds. **C**, The period of suprathreshold activity near 50 seconds is binned at 4 ms, showing that activity is not actually synchronous at higher temporal resolution. Activity here spans three bins and is preceded and terminated by bins with no activity. **D**, The activity shown in B is presented as a spatio-temporal pattern on the multielectrode array grid. In this case, a pattern of three frames is shown. **E**, Six cases of spatio-temporal activity patterns are shown that were significantly repeating in a one hour period. Here active electrodes are shown as darkened squares on the electrode grid, where darker squares indicate larger amplitude signals and lighter squares indicate smaller amplitudes. Next to each pair of patterns is the time, in minutes, between observations of the patterns. Since the cultures were grown in isolation from sensory inputs, these results indicate that reproducible activity patterns can be generated by cortical circuits through self-organizing mechanisms. Figures adapted from Beggs and Plenz, 2003, 2004

these patterns suggests that Hebbian rules and recurrent connectivity may be sufficient conditions for stable activity patterns.

So far, these findings seem consistent with the simple attractor neural network model described previously. But does activity in these networks show evidence of becoming progressively more like a stored pattern? Is the network configuration being drawn in to an attractor? Interestingly, very few laboratories sought to examine the dynamics of activity patterns in these systems. Because of this, the attractor hypothesis in its fullest form was not truly evaluated by the work described above.

Recently, Wills and colleagues (Wills TJ et al., 2005) have made progress on this issue with an ingenious set of experiments performed in awake, behaving rats. They implanted multiple electrodes in the hippocampus and then placed rats in an arena with a base that could be progressively morphed from a circle to a square. Consistent with previous studies, they found that a particular activity pattern of firing in hippocampal neurons occurred when the rat was in the circular arena, and that this pattern was different from the pattern that occurred when the rat was placed in the square arena. After testing that these representations were stable, they then changed the shape of the arena to be like that of a square with rounded edges, intermediate between a circle and a square. When the rat was placed in this new hybrid arena, the activity pattern that initially appeared on the electrodes was not like that seen from the circular or the square arena. Over two minutes, though, the activity pattern progressively became more like either that seen from the circular arena or that seen from the square arena. This is exactly what would be expected if the network state were being drawn into an attractor. They also showed that slight morphs away from the circular shape usually resulted in network activity becoming like the pattern seen from the purely circular arena; similar effects were shown for slight morphs away from the square shape. These data were consistent with the basin of attraction seen in the simple network model presented earlier.

Although the results of this impressive experiment qualitatively agreed with all of the major features of the attractor network model, several areas still remained to be explored. It was not clear if the dynamics seen in this system was caused by the circuitry within the hippocampus or by the cooperative action of other brain areas that projected to the hippocampus. With convergence to an attractor state taking about two minutes, it seemed likely that other brain areas were involved. It would also be desirable to go beyond a qualitative description and to quantify the dynamics more precisely.

4 Tools for Quantifying Local Network Dynamics

How can the dynamics of neural networks be quantified? Fortunately methods from dynamical systems theory have been developed and these have successfully been applied to electronic circuits (Huberman BA et al., 1980), driven

pendulums (Baker GL and JP Gollub , 1996), chemical reactions (Kilgore MH et al., 1981) and a host of other phenomena (Crutchfield JP et al., 1986; Nicolis G and I Prigogine, 1989). With some changes, these methods can also be used to describe both simulated and living neural networks. In this section, we will briefly describe some of these tools and note how they could be used to sharpen the description of network dynamics that was qualitatively outlined in the previous section. For a more detailed treatment of this topic, the reader is referred to the chapter by Jirsa and Breakspear in this handbook.

We will assume that we wish to describe the dynamics of a network composed of m neurons. Let x_i represent a variable of interest, for example, the voltage, of neuron i. The configuration of activity in the network at time t can then specify a location X^t in m-dimensional state space:

$$X^t = (x_1^t, x_2^t, x_3^t, \ldots x_m^t) \ .$$

In these coordinates, we can plot network activity at times $t + 1$, $t + 2$, $t+3 \ldots t+n$ and we can construct a trajectory (also called an orbit) by linking these locations in state space (also called phase space) as shown in Fig. 3.

Describing the dynamics of the network amounts to describing how trajectories evolve over time. Recall that in the attractor network model, the network state will evolve toward a stored configuration. Trajectories starting within the same basin of attraction will therefore tend to flow toward each other over time, minimizing the distance between them. So to explore the attractor network hypothesis, we will need to quantify distances in state space.

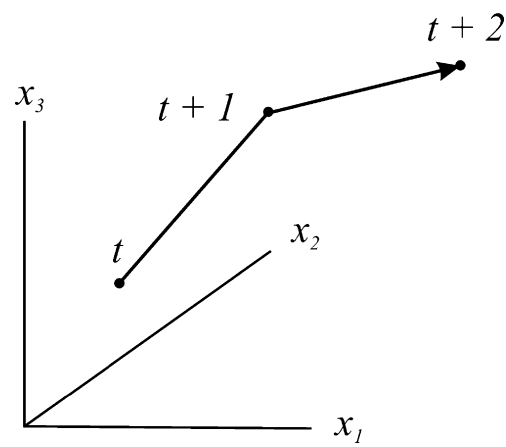


Fig. 3. A trajectory in state space. Three axes, x_1, x_2, and x_3 are shown, which could represent the states of three neurons. By plotting the values of the network state variables (x_1, x_2, x_3) at times t, $t + 1$, and $t + 2$, a succession of positions can be linked to form a trajectory through state space. The trajectory is shown here as a bent arrow

Distances between points X^t and Y^t can be measured by some metric, like the Euclidean distance:

$$d_{X^t Y^t} = \sqrt{(y_1^t - x_1^t)^2 + (y_2^t - x_2^t)^2 + (y_3^t - x_3^t)^2 + \ldots (y_m^t - x_m^t)^2}.$$

Other metrics are also suitable. For example, the Hamming distance, which is just the number of digits that are different between two binary numbers (e.g., [1 0 1] and [0 0 1] have a Hamming distance of 1), could be used for a network with only binary neurons. The rate of growth in distance between two initially close trajectories can be quantified by the Lyapunov exponent λ (Wolf A et al., 1985), which is related to the distance between trajectories at two points in time. This is illustrated in Fig. 4, where trajectories begin from two points that are close together in state space. The distance between these two starting points is measured as d_{start}. The network is allowed to evolve over time from each point, causing two trajectories to be traced out in state space. After a time T, the distance between two points on the trajectories is measured as d_{finish}. The Lyapunov exponent in bits/sec is then given by:

$$\lambda = \frac{1}{T} \log_2 \left(\frac{d_{finish}}{d_{start}} \right).$$

In practice it is good to keep T small so that λ will closely approximate the instantaneous divergence between the trajectories. By manipulating this equation, we can more clearly see how λ describes the exponential rate at which two trajectories separate in state space after T time steps:

$$d_{start} \cdot 2^{\lambda T} \cong d_{finish}.$$

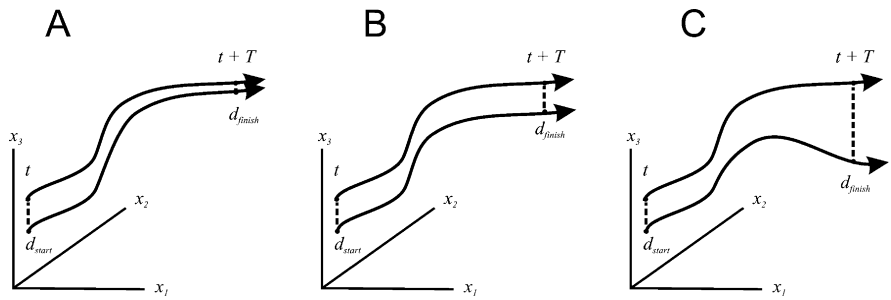


Fig. 4. The Lyapunov exponent quantifies dynamics. **A**, Converging trajectories. Two trajectories in state space, shown as curved lines with arrowheads, are separated by a distance d_{start} at time t. At time $t + T$, they are separated by a distance of d_{finish}. The ratio (d_{finish}/d_{start}) can be used to determine whether or not the trajectories are flowing together over time. In this case, they become closer over time, indicating attractive dynamics. **B**, Parallel trajectories. Here, the ratio (d_{finish}/d_{start}) is one, indicating neutral dynamics. **C**, Diverging trajectories. Here the ratio (d_{finish}/d_{start}) is greater than one, indicating chaotic dynamics

The Lyapunov approach to quantifying discrete network dynamics has been developed by Derrida and colleagues (Derrida B and Y Pomeau, 1986; Derrida B and G Weisbuch, 1986), as well as by others (Bertschinger N and T Natschlager, 2004). This method is especially useful when working with simulated networks, where it is easy to start a network from a particular point in state space by just specifying the values of all the state variables. The simulation can then be run for T time steps to produce a trajectory. It is also easy to produce a trajectory from a nearby point in state space and to measure the resulting distances between trajectories. For living neural networks, however, this approach is more difficult to implement, as it is presently impossible to specify all the state variables at a given time. Electrical stimulation can overcome this to some extent by causing a subset of neurons to all be active at the same time; trajectories after stimulation can then be measured. But background activity can not be completely controlled, and this has been found to play a large role in determining network responses to stimulation (Arieli A et al., 1996).

There are three general types of dynamics that can be identified with this method. Attractive dynamics is characterized by $\lambda < 0$, causing nearby trajectories to become closer over time (Fig. 4A). Systems dominated by attractive dynamics are very stable, and have one or more basins of attraction. In the attractor model that we previously described, these basins would lead to attractor states that could represent configurations stored in long-term memory. However, these networks are so stable that it is difficult to control their trajectories and steer them away from attractors. Perturbations to the network are mostly ineffective at changing the state that it settles into. Neutral dynamics is characterized by $\lambda \approx 0$, causing nearby trajectories to preserve distance over time (Fig. 4B). Here, perturbations to the network produce commensurate changes in output. Systems with predominantly neutral dynamics are therefore marginally stable, meaning that trajectories will largely persist in their given course under mild perturbations. With the appropriate inputs, it is possible to control trajectories in networks with neutral dynamics. Chaotic dynamics is characterized by $\lambda > 0$ causing nearby trajectories to become more separated over time (Fig. 4C). Small perturbations are amplified, making these networks intrinsically unstable and difficult, but not impossible, to control (Ding M et al., 1996; Ditto WL and K Showalter, 1997).

The Lyapunov exponent can be used to describe trajectories in all regions of state space that are visited by the network. However, just because one region of state space shows attractive dynamics does not necessarily mean that all other regions will also. Figure 5 shows that state space can contain a variety of features: fixed points, saddle points, and limit cycles. The trajectories passing through fixed points are all either leading into the point or leading away from it. If they are leading into the point, then there is a stable fixed point; if they are leading away from it, then there is an unstable fixed point. The attractor network model discussed previously uses stable fixed points to encode long-term memories. Unstable fixed points are repulsive to trajectories

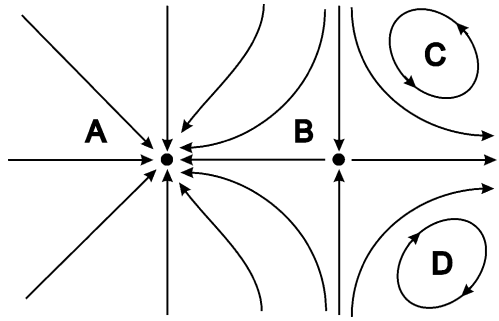


Fig. 5. Example features of state space. **A**, A stable fixed point has only trajectories leading into it. **B**, A saddle point has trajectories leading into it along one axis, but leading out of it along another axis. **C**, **D**, Limit cycles are trajectories that are closed on themselves

and are rarely visited by the system. Saddle points are both attractive and repulsive, depending on the direction from which they are approached. They have trajectories that lead into them from one direction, and trajectories that lead away from them in another direction. Trajectories will often move toward a saddle point, only to be repulsed from it when they get too close. They then may be drawn toward another saddle point and repulsed again, displaying itinerant behavior as they visit different saddle points (Rabinovich M et al., 2001). Limit cycles occur when the network continually oscillates in state space, as represented by trajectories that form closed loops with themselves.

To fully characterize the dynamics, one would have to map the entire state space of the system. This is in practice impossible, so most experiments report only the dynamics seen in a small subset of the state space. Fortunately, characterizing the dynamics in some reduced dimensional space is often good enough to get an approximate picture of the dynamical system as a whole. The process of knowing how to reduce dimensionality and which variables may be omitted is beyond the scope of this chapter. The reader is referred to (Abarbanel HD and MI Rabinovich, 2001; Abarbanel MDI, 1996; Kantz H and T Schreiber, 2004; Strogatz SH, 1994) for more detailed discussions on this topic.

5 How Connectivity Influences Local Network Dynamics

With the methods described above, we can now examine how network connectivity influences dynamics. Since it is difficult to manipulate connectivity in living neural networks, we will only discuss here results from computational models. In what follows, we will introduce a simple model with tunable connectivity. We will show that this model qualitatively captures the main features of network activity observed in some experiments. We will then manipulate the connectivity of the model to explore its effects on dynamics.

Consider a network model with N neurons or processing units, each allowed to be in only one of two states, either active (1) or inactive (0). To allow for complete generality, let us connect each unit to every other unit (Fig. 6A). We can later control the strengths of these connections, setting some of them to zero, so as to sculpt the connectivity of the network. Units can become active in one of two ways, either through spontaneous or driven activity. Each unit will have some small probability, p_{spont}, of being spontaneously active at a given time step. A unit may also become active if it is driven by another active unit that makes a connection with it. If a unit is not spontaneously active or driven at a given time step, it will be inactive.

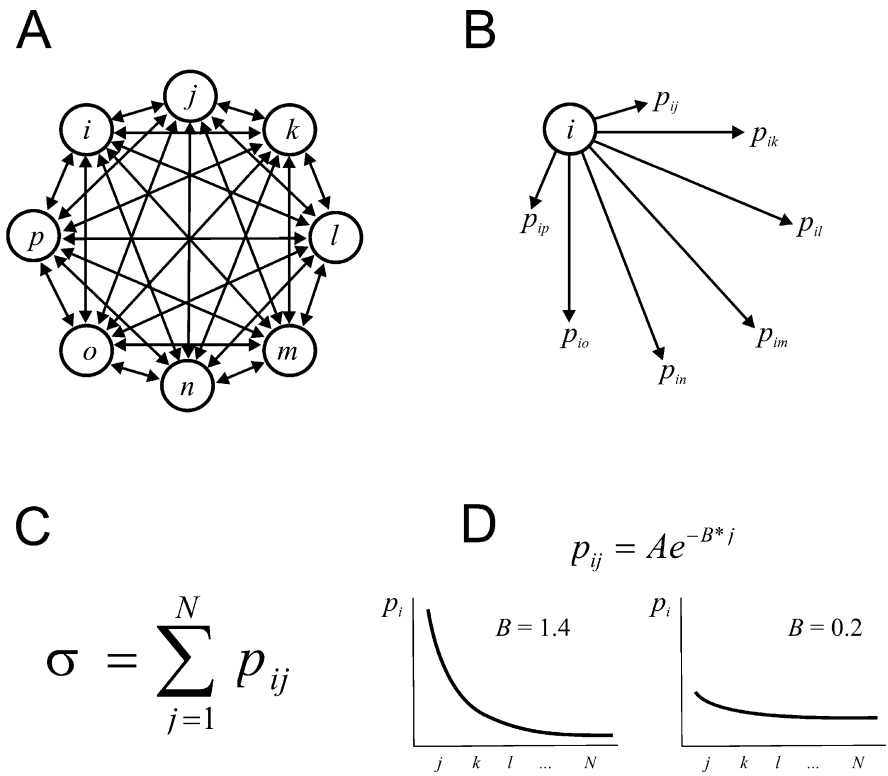


Fig. 6. Network model with tunable connectivity. **A**, Network initially has all-to-all connectivity, but selected connection strengths can be set to zero. A network with $N = 8$ units is shown. **B**, Each unit i has a set of transmission probabilities: $\{p_{ij}, p_{ik}, \ldots p_{iN}\}$ that determines connection strengths. **C**, The sum of the transmission probabilities emanating from a given unit i will determine the branching parameter σ for that unit. **D**, The distribution of transmission probabilities can be made sharp or flat by adjusting the exponent B. The normalization constant, A, makes the probabilities sum to σ. As discussed in the text, tuning the branching parameter σ or the distribution exponent B can influence network dynamics

Activity can propagate from unit i to unit j through a connection that has a transmission probability p_{ij} that is constrained to be between zero and one (Fig. 6B). Transmission is simple and works like this: If a unit i is active, then unit j will become active in the next time step if a randomly drawn number is less than the transmission probability p_{ij}. In other words, unit i will transmit to unit j with probability p_{ij}. Unlike traditional integrate-and-fire neuron models, these units do not sum all of the incoming activity and then fire if this sum is over a threshold. They simply fire if one of the units connected to them is active and if transmission between them is successful. Given this arrangement, activity in the model typically originates spontaneously at one or a few units and then propagates through connections to other units in the network. While this model may seem too simplistic, it actually does a good job of reproducing phenomena observed in the data, as will be explained more below. If a parsimonious model can successfully capture the main features of the data, then this suggests that network dynamics may be governed by a few simple principles (Haldeman C and JM Beggs, 2005).

The connectivity may be tuned in one of two ways. First, the sum of the transmission probabilities emanating from each unit may be scaled from 0 (where each $p_{ij} = 0$) to N (where each $p_{ij} = 1$). Let us define the branching parameter, σ, as this sum:

$$\sigma \equiv \sum_{j=1}^{N} p_{ij}.$$

The branching parameter will serve to scale all connection strengths (Fig. 6C). Second, the distribution of transmission probabilities from each unit may be controlled from sharp (where only one connection has a transmission probability of 1 and all the other connections are 0) to homogeneous (where all connection strengths have equal transmission probabilities equal to $1/N$)(Fig. 6D). There are many different types of distributions that could be used here, but for simplicity we will only consider distributions that are defined by an exponential function:

$$p_{ij} \equiv Ae^{-B*j},$$

where A is a scaling constant that keeps the sum of the transmission probabilities equal to σ, and B is the exponent that determines how sharp (B large) or flat (B small) the distribution will be.

How well can a simple model like this capture features from actual data? In experiments with organotypic cortical cultures, Beggs and Plenz (Beggs JM and D Plenz, 2003) found that suprathreshold local field potential activity (Fig. 2A) at one electrode was, on average, followed by activity in one other electrode in the next time step. When the model is tuned to have a branching parameter $\sigma = 1.0$, it reproduces this result faithfully. This should not be too surprising, though, since it is well known that for a branching process, σ gives the expected number of descendants from a single ancestor (Harris TE, 1989). What is somewhat less expected is that the distribution of "avalanche" sizes

produced by the model also closely matches the distribution from the data. Here, the avalanche size is just the total number of electrodes activated in one spatio-temporal pattern of activity. Representative patterns of activity are shown in Fig. 2D and 7B, and consist of consecutively active frames that are bracketed in time by inactive frames. When the probability of an avalanche is plotted against its size, the result is a power law, as shown in Fig. 7A. Power law distributions are often found in complex systems and can be used to describe domain sizes in magnets during a phase transition (Stanley HE,

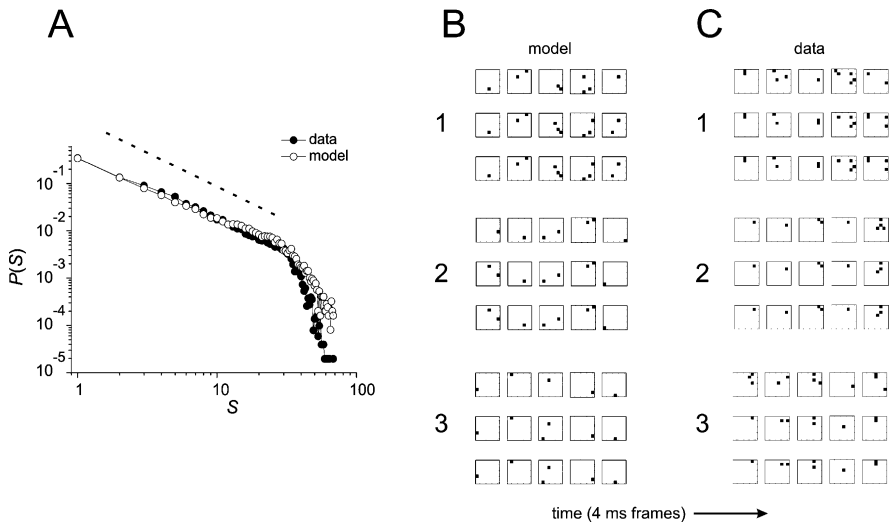


Fig. 7. The model captures important features of the data. **A**, When the probability of occurrence is plotted against avalanche size, a nearly straight line is formed in a log-log graph. This line indicates that the relationship between probability and size can be described by a power law: $P(S) = S^{-\alpha}$, where P is the probability of observing an avalanche of size S, and S is the total number of electrodes activated in the avalanche. For a critical branching process, the exponent of the power law, α, is predicted to be $-3/2$ (dashed line). Filled circles are from 5 hours of spontaneous activity in an acute slice, while open circles show results from the model when the branching parameter $\sigma = 1$. Note that the power law begins to cut off near $S = 35$, since the number of electrodes in the array is 60. **B**, Reproducible patterns of activity generated by the model. Each large white square represents the pattern of active electrodes on the array at a given 4 ms time step. Active electrodes are shown as small black squares. Patterns shown are all five time steps long. Note that patterns within groups 1 through 3 are not exactly the same, even though all groups were statistically significant. **C**, Reproducible patterns generated by acute cortical slices. Note general similarity to patterns produced by the model. Data from acute cortical slices are generally similar to data produced by organotypic cortical cultures (compare to patterns shown in Fig. 2D), suggesting common principles of operation. Because the model reproduces general features of the data, it may serve as a useful tool for exploring links between connectivity and dynamics

1987; Yeomans JM, 1992), forest fire sizes (Malamud BD et al., 1998), earthquake magnitudes (Gutenberg B and CF Richter, 1941) and sizes of simulated sand pile avalanches (Paczuski M et al., 1996). Since the sizes of activity patterns from cortical cultures also fit a power law, we called them "neuronal avalanches." Power law distributions also suggest, but do not prove, that a system is operating near a critical point (Bak P, 1996; Jensen HJ, 1998). The power law is a consequence of the branching parameter being close to unity. When $\sigma = 1$, activity propagates in a nearly sustained manner but eventually dies out because transmission is stochastic. Another aspect of the data that can be reproduced by this simple model is the reproducible activity patterns themselves. As shown in Fig. 7B, the patterns produced by the model are qualitatively similar to those produced by cortical slices (see Fig. 2D for similar patterns produced by cultures). These patterns are caused by inequalities in the connection strengths of the model. Although each transmission is probabilistic, there will be some preferred patterns of activity in the network because some connections are stronger than others. Because this parsimonious model qualitatively reproduces two main features from living network data, it seems plausible that we could use the model to predict how connectivity will influence dynamics in real neural networks.

How is dynamics in the model affected by changes in the branching parameter σ? To explore this, we can tune all units in the network to have a given σ. We then let the network evolve over time from many pairs of closely spaced starting configurations. By measuring the distances between trajectories from thousands of pairs of configurations, we can estimate the Lyapunov exponent λ for the network. For $\sigma < 1$, transmission probabilities are weak and avalanches tend to die out because the average number of descendants from a given ancestor is less than one. This causes trajectories to become more similar over time, since fewer and fewer units are active and distances decrease. In this case, the dynamics is predominantly attractive and $\lambda < 0$. For $\sigma \approx 1$, connections between units are stronger and activity is nearly sustained since the average number of descendants from a given ancestor is one. Here, distances between nearby trajectories are preserved and $\lambda \approx 0$, indicating neutral dynamics. For $\sigma > 1$, the number of active units in the network increases with every time step, causing slight distances in state space to be amplified. As a result, trajectories tend to diverge in state space. The Lyapunov exponent is $\lambda > 0$, indicating chaotic dynamics with its typical sensitive dependence on initial conditions. These results clearly suggest that the *sum* of connection strengths, or weights, can determine the dynamical regime of a network (Fig. 8A).

But what happens if we begin to change the *distribution* of weights coming from each unit? In simulated neural networks, this question was pursued by Bertshinger and Natschlager (Bertschinger N and T Natschlager, 2004), who found that dynamics could be tuned by changing the variance of a Gaussian weight distribution. They showed that small variances led to attractive dynamics while large variances led to chaotic dynamics. Inspired

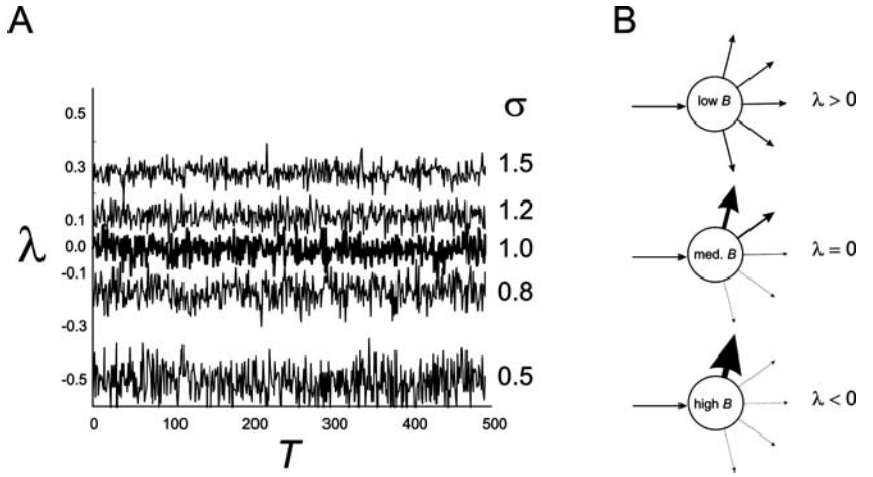


Fig. 8. How connectivity influences network dynamics. **A**, The Lyapunov exponent λ is plotted for each of 'T' time steps for different values of the branching parameter σ. The branching parameter governs the sum of transmission probabilities from each unit of the network. When σ is close to the critical value of 1, dynamics is neutral, and λ hovers around 0. As σ is increased, dynamics becomes chaotic ($\lambda > 0$); as σ is decreased, dynamics becomes attractive ($\lambda < 0$). **B**, The distribution of transmission probabilities emanating from each unit also influences dynamics. Three different types of units are shown, representing the three different types of exponential distributions that were examined. Thick arrows represent high transmission probabilities. The top unit shows transmission probabilities when the exponent B is low and the distribution is homogeneous. In this case, each unit acts to disperse trajectories, causing chaotic dynamics. The middle unit corresponds to intermediate values of B, where one to two transmission probabilities are strong. Here, each unit acts to focus trajectories, but with some dispersion, causing neutral dynamics. The lower unit illustrates the highly skewed distribution caused by large values of B. Here one connection dominates and all the rest are essentially zero. Units with high B distributions act to focus trajectories, leading to attractive dynamics. Figure 8 A is modified from Haldeman and Beggs, 2005, copyright American Physical Society, and is reproduced with permission

by their approach, we here use an exponential distribution whose sharpness can be tuned through an exponent B, and we explore how B affects λ. In these simulations, we use a network with 64 units that has 8 connections per unit. Qualitatively similar results obtain for networks with 64 connections per unit, suggesting that these findings are quite general. For small B ($0 \leq B < 1.0$), distributions are nearly flat and each connection has roughly the same probability of transmission. In this case, activity coming in to a unit will be spread widely and randomly to other connected units. This tends to disperse trajectories and leads to chaotic dynamics where $\lambda > 0$. For intermediate values of $B(1.2 \leq B < 1.8)$, one or two connections have transmission probabilities that are much larger than all the rest. Here activity coming in to

a unit will tend to be transmitted to only one or two other units. This leads to propagation in which there is a balance of spreading and focus. While there is some variability in the paths that trajectories take, there is one path that is traveled most of the time. On average, dynamics tend to be neutral and $\lambda \approx 0$. For large values of $B(1.8 < B)$, one of the transmission probabilities is very near 1, while all of the others are near zero. So while a unit may receive convergent activation from two other units in the previous time step, it will almost always activate only one unit in the next time step. Under these conditions, units serve to bring different trajectories together, thus reducing distances over time and causing attractive dynamics with $\lambda < 0$. Together, these simulations show that the distribution of connection strengths can also set the dynamics of a network (Fig. 8B).

How do changes in the number of connections affect dynamics? Although not directly in the field of neural networks, Stuart Kauffman has pursued this question in network models of gene regulatory networks. Since his studies are very likely to be relevant to our topic, we briefly mention them here. Kauffman and colleagues (Kauffman S, 1969; Kauffman S et al., 2003; Kauffman SA and S Johnsen, 1991) examine networks where each binary unit can be either on (1) or off (0), and where each unit performs some Boolean function (e.g., AND, OR) on its inputs. Units are connected randomly, and the number of connections into each unit is determined by an order parameter K. Kauffman shows in these random Boolean networks that when $K > 3$, trajectories are very sensitive to small perturbations and dynamics is chaotic. When $K = 2$, however, trajectories are stable with respect to perturbations and the networks appear to operate at a critical point (Bornholdt S and T Rohlf, 2000). For $K < 2$, nearly all trajectories quickly fall into attractors. Kauffman and others (Gutowitz H and C Langton, 1995) have suggested that K governs a phase transition in these networks as it controls their dynamics. In some ways, high K networks may be similar to the neural network model described above when the distribution exponent B is small and all transmission probabilities are nearly equal. For intermediate values of B, one or two transmission probabilities are strong, and this may correspond to the critical case where $K = 2$ in Kauffman's networks. These possible connections are intriguing and deserve further exploration.

6 Discussion and Prospects

But why should dynamics matter? The dynamical regime of a network can strongly influence the types of computations it is able to perform (Vogels TP et al., 2005). Many models and experiments suggest that local networks support attractive dynamics (Amit Y and M Mascaro, 2001; Brunel N, 2000; Hopfield JJ, 1982; Jin DZ, 2002; Seung HS, 1998; Wills TJ et al., 2005). As mentioned earlier, strongly attractive dynamics is naturally good for setting up attractor states in which long-term memories can be stably stored. Such

dynamics is also desirable for pattern completion, since a fragment of a stored pattern can be used as a cue to get the network into a state where it is near a basin of attraction and likely to evolve into the stored memory configuration. Moreover, attractive dynamics supports computations that favor categorization since they cause different stimuli to be grouped into the same response. For example, if a Wolfhound, a Chihuahua and a Beagle were all represented by positions in state space, attractive dynamics could cause trajectories from these points to all flow together, making it easy to set up the category of "dog." But the stability conferred by attractive dynamics also makes it difficult to steer trajectories away from strong attractors. Networks dominated by attractive dynamics would seem to lack flexibility.

In contrast, chaotic dynamics supports computations that favor discrimination since subtle differences in stimuli can produce widely different responses. Here too, there are a number of models and experiments that suggest that chaotic dynamics are prevalent in the brain (Aitken PG et al., 1995; Babloyantz A and A Destexhe, 1986; Breakspear M et al., 2003; Freeman WJ, 1994; Schiff SJ et al., 1994; van Vreeswijk C and H Sompolinsky, 1996). This dynamics could be useful in sensory systems where there is a great need to notice details of the incoming information stream. For example, whether a rabbit stays and eats or rapidly flees may be determined by only a few blades of grass in the visual field that seem to be moving in an unusual way. There have also been proposals that chaotic processing units could be used to perform logical or arithmetic computations since such units are naturally nonlinear (Sinha S and WL Ditto, 1999). However, networks with trajectories that rapidly diverge are unstable unless they are controlled.

With neutral dynamics, differences in inputs produce commensurate differences in responses. Not surprisingly, there are models and experiments that suggest this type of dynamics is used too (Beggs JM and D Plenz, 2003; Bertschinger N and T Natschlager, 2004; Haldeman C and JM Beggs, 2005; Latham PE and S Nirenberg, 2004; Maass W et al., 2002). This dynamics supports computations that favor efficient information transmission since a one-to-one mapping between stimuli and responses is maintained. They may also be optimal for information storage (Beggs JM and D Plenz, 2004; Haldeman C and JM Beggs, 2005). Several researchers have pointed out that neutral dynamics, "at the edge of chaos," may also be best for performing the widest variety of computations because it combines some of the variety of chaos with some of the stability of attractive systems (Bertschinger N and T Natschlager, 2004; Beggs 2007). It is argued that useful computations require both nonlinear transformations and stable representations of information. Perhaps neocortex, which is essential for higher-level computations, has largely neutral dynamics (Maass W et al., 2002; Natschlager T and W Maass, 2005).

To advance research in this area it will be necessary to form a tighter link between models and experiments. Many of the ideas about how connectivity influences dynamics described above have not yet been tested in living neural networks. Since nature often defies our expectations, it is essential that we

develop better ways of interrogating networks of neurons. With advances in technology in the next ten years (Frechette ES et al., 2005), it may be possible to stimulate and record from thousands of neurons for periods of weeks at a time. The huge data sets that are likely to be produced will hopefully allow us to map the state space of living neural networks more closely.

It will also be important to investigate how different network topologies (e.g., random, small-world, scale-free) explicitly influence dynamics. The simulations described above treated all nodes in the network equivalently, but this is certainly a simplification. What happens when some nodes have different branching parameters and transmission probabilities than others? What if some nodes have more connections than others? These issues are only now beginning to be explored (Fox JJ and CC Hill, 2001), as the network topology of the brain at the local network level (Netoff TI et al., 2004; Song S et al., 2005) and at the large scale level (Achard S et al., 2006; Eguiluz VM et al., 2005; Sporns O et al., 2005; Sporns O and JD Zwi, 2004; Stam CJ et al., 2006) is still not well known. The connectivity patterns, and therefore the dynamics, at these different levels may not necessarily be the same (Breakspear M and CJ Stam, 2005; Jirsa VK, 2004).

Another area that deserves much attention is the relationship between dynamics and connectivity: How does brain activity, both acutely and chronically, alter the connectivity of neural networks? While activity-dependent synaptic plasticity has been extensively studied, most of this work has centered on how stimulation at one or a few synapses influences synaptic efficacy. There is a need to expand the focus to explore how activity at the local network level may influence synaptic plasticity. In vivo, transmission at a single synapse is embedded in the context of rich background activity that is very influential (Leger JF et al., 2005). From this perspective, functional connectivity is very dynamic and may be different from the underlying structural connectivity (Sporns O et al., 2000). Since it has been shown that large-scale network connectivity can change from wakefulness to sleep (Massimini M et al., 2005), it seems likely that it would also change during transitions to other brain states as well, like seizures. Similar changes at the local network level should also be investigated. While it may be difficult to disentangle the contributions of connectivity and dynamics in these situations, their complexity suggests that these situations will be interesting and fruitful areas for further research.

In the previous sections we have shown how early models of memory storage in local recurrent networks led many to search for attractors in neurophysiological data. While numerous examples of reproducible activity patterns in living neural networks have been found, very few experimental studies have addressed the dynamics of these networks quantitatively. By measuring the Lyapunov exponent in simple network models, it has become clear that network connectivity can profoundly influence dynamics. Experimental work in the future will hopefully begin to quantitatively address the dynamics of local cortical networks, perhaps even revealing how trajectories in cortical columns perform computations that form the building blocks of cognition.

Acknowledgements

This work was supported by the National Science Foundation and Indiana University.

References

Abarbanel HD, Rabinovich MI (2001) Neurodynamics: nonlinear dynamics and neurobiology. Current Opinion in Neurobiology 11: 423–430

Abarbanel HDI (1996) Analysis of observed chaotic data. New York: Springer

Abeles M, Bergman H, Margalit E, Vaadia E (1993) Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. J Neurophysiol 70: 1629–1638

Achard S, Salvador R, Whitcher B, Suckling J, Bullmore E (2006) A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. J Neurosci 26: 63–72

Aitken PG, Sauer T, Schiff SJ (1995) Looking for chaos in brain slices. J Neurosci Methods 59: 41–48

Amit DJ (1989) Modeling brain function : the world of attractor neural networks. Cambridge ; New York: Cambridge University Press

Amit Y, Mascaro M (2001) Attractor networks for shape recognition. Neural Comput 13: 1415–1442

Anderson JA, Silverstein JW, Ritz SA, Jones RS (1977) Distinctive Features, Categorical Perception, and Probability Learning: Some Applications of a Neural Model. Psychological Review 84: 413–451

Arieli A, Sterkin A, Grinvald A, Aertsen A (1996) Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. Science 273: 1868–1871

Babloyantz A, Destexhe A (1986) Low-dimensional chaos in an instance of epilepsy. Proc Natl Acad Sci U S A 83: 3513–3517

Bak P (1996) How nature works : the science of self-organized criticality. New York, NY, USA: Copernicus

Baker GL, Gollub JP (1996) Chaotic dynamics : an introduction. Cambridge; New York: Cambridge University Press

Baker SN, Lemon RN (2000) Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance levels. J Neurophysiol 84: 1770–1780

Beggs JM, Plenz D (2003) Neuronal avalanches in neocortical circuits. J Neurosci 23: 11167–11177

Beggs JM, Plenz D (2004) Neuronal avalanches are diverse and precise activity patterns that are stable for many hours in cortical slice cultures. J Neurosci 24: 5216–5229

Beggs JM (2007) The criticality hypothesis: How local cortical networks might optimize information processing. Proceedings of the Philosophical Society A (submitted)

Ben-Shaul Y, Drori R, Asher I, Stark E, Nadasdy Z, Abeles M (2004) Neuronal activity in motor cortical areas reflects the sequential context of movement. J Neurophysiol 91: 1748–1762

Bertschinger N, Natschlager T (2004) Real-time computation at the edge of chaos in recurrent neural networks. Neural Comput 16: 1413–1436

Bornholdt S, Rohlf T (2000) Topological evolution of dynamical networks: global criticality from local dynamics. Phys Rev Lett 84: 6114–6117

Breakspear M, Stam CJ (2005) Dynamics of a neural system with a multiscale architecture. Philos Trans R Soc Lond B Biol Sci 360: 1051–1074

Breakspear M, Terry JR, Friston KJ (2003) Modulation of excitatory synaptic coupling facilitates synchronization and complex dynamics in a biophysical model of neuronal dynamics. Network 14: 703–732

Brown EN, Frank LM, Tang D, Quirk MC, Wilson MA (1998) A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. J Neurosci 18: 7411–7425

Brunel N (2000) Dynamics of networks of randomly connected excitatory and inhibitory spiking neurons. J Physiol Paris 94: 445–463

Canli T, Desmond JE, Zhao Z, Gabrieli JD (2002) Sex differences in the neural basis of emotional memories. Proc Natl Acad Sci U S A 99: 10789–10794

Chi Z, Margoliash D (2001) Temporal precision and temporal drift in brain and behavior of zebra finch song. Neuron 32: 899–910

Cohen MA, Grossberg S (1983) Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. IEEE Transactions on Systems, Man, and Cybernetics SMC-13: 815–826

Cossart R, Aronov D, Yuste R (2003) Attractor dynamics of network UP states in the neocortex. Nature 423: 283–288

Crutchfield JP, Farmer JD, Packard NH, Shaw RS (1986) Chaos. Sci Am 255: 46–&.

Dave AS, Margoliash D (2000) Song replay during sleep and computational rules for sensorimotor vocal learning. Science 290: 812–816

Deregnaucourt S, Mitra PP, Feher O, Pytte C, Tchernichovski O (2005) How sleep affects the developmental learning of bird song. Nature 433: 710–716

Derrida B, Pomeau Y (1986) Random Networks of Automata - a Simple Annealed Approximation. Europhys Lett 1: 45–49

Derrida B, Weisbuch G (1986) Evolution of Overlaps between Configurations in Random Boolean Networks. J Phys-Paris 47: 1297–1303

Ding M, Yang W, In VV, Ditto WL, Spano ML, Gluckman B (1996) Controlling chaos in high dimensions: Theory and experiment. Physical Review E Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics 53: 4334–4344

Ditto WL, Showalter K (1997) Introduction: Control and synchronization of chaos. Chaos 7: 509–511

Eguiluz VM, Chialvo DR, Cecchi GA, Baliki M, Apkarian AV (2005) Scale-free brain functional networks. Phys Rev Lett 94: 018102

Fletcher P, Buchel C, Josephs O, Friston K, Dolan R (1999) Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging. Cereb Cortex 9: 168–178

Foss J, Longtin A, Mensour B and Milton JG (1996) Multistability and delayed recurrent loops. Phys. Rev. Lett. 76: 708–711

Foss J and Milton J (2000) Multistability in recurrent neural loops arising from delay. J. Neurophysiol. 84: 975–985

Fox JJ, Hill CC (2001) From topology to dynamics in biochemical networks. Chaos 11: 809–815

Frechette ES, Sher A, Grivich MI, Petrusca D, Litke AM, Chichilnisky EJ (2005) Fidelity of the ensemble code for visual motion in primate retina. J Neurophysiol 94: 119–135

Freeman WJ (1994) Neural networks and chaos. J Theor Biol 171: 13–18

Gutenberg B, Richter CF (1941) Seismicity of the earth. [New York]: The Society

Gutowitz H, Langton C (1995) Mean field theory of the edge of chaos. Advances in Artificial Life 929: 52–64

Hahnloser RH, Kozhevnikov AA, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature 419: 65–70

Haldeman C, Beggs JM (2005) Critical branching captures activity in living neural networks and maximizes the number of metastable States. Phys Rev Lett 94: 058101

Harris TE (1989) The theory of branching processes. New York: Dover Publications

Hebb DO (1949) The organization of behavior; a neuropsychological theory. New York,: Wiley

Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A 79: 2554–2558

Hopfield JJ (1984) Neurons with graded response have collective computational properties like those of two-state neurons. Proc Natl Acad Sci U S A 81: 3088–3092

Hopfield JJ, Tank DW (1986) Computing with neural circuits: a model. Science 233: 625–633

Huberman BA, Crutchfield JP, Packard NH (1980) Noise Phenomena in Josephson Junctions. Appl Phys Lett 37: 750–752

Jantzen KJ, Steinberg FL, Kelso JA (2005) Functional MRI reveals the existence of modality and coordination-dependent timing networks. Neuroimage 25: 1031–1042

Jensen HJ (1998) Self-organized criticality : emergent complex behavior in physical and biological systems. Cambridge, U.K. ; New York: Cambridge University Press

Jin DZ (2002) Fast convergence of spike sequences to periodic patterns in recurrent networks. Phys Rev Lett 89: 208102

Jirsa VK (2004) Connectivity and dynamics of neural information processing. Neuroinformatics 2: 183–204

Kantz H, Schreiber T (2004) Nonlinear time series analysis. Cambridge, UK ; New York: Cambridge University Press

Kauffman S (1969) Homeostasis and Differentiation in Random Genetic Control Networks. Nature 224: 177

Kauffman S, Peterson C, Samuelsson B, Troein C (2003) Random Boolean network models and the yeast transcriptional network. P Natl Acad Sci USA 100: 14796–14799

Kauffman SA, Johnsen S (1991) Coevolution to the Edge of Chaos - Coupled Fitness Landscapes, Poised States, and Coevolutionary Avalanches. J Theor Biol 149: 467–505

Kelso SR, Ganong AH, Brown TH (1986) Hebbian synapses in hippocampus. Proc Natl Acad Sci U S A 83: 5326–5330

Kenet T, Bibitchkov D, Tsodyks M, Grinvald A, Arieli A (2003) Spontaneously emerging cortical representations of visual attributes. Nature 425: 954–956

Kilgore MH, Turner JS, Mccormick WD, Swinney H (1981) Periodic and Chaotic Oscillations in the Belousovzhabotinskii Reaction. B Am Phys Soc 26: 362–362

Kirkwood A, Bear MF (1994) Hebbian synapses in visual cortex. J Neurosci 14: 1634–1645

Latham PE, Nirenberg S (2004) Computing and stability in cortical networks. Neural Comput 16: 1385–1412

Lee AK, Wilson MA (2002) Memory of sequential experience in the hippocampus during slow wave sleep. Neuron 36: 1183–1194

Lee AK, Wilson MA (2004) A combinatorial method for analyzing sequential firing patterns involving an arbitrary number of neurons based on relative time order. J Neurophysiol 92: 2555–2573

Leger JF, Stern EA, Aertsen A, Heck D (2005) Synaptic integration in rat frontal cortex shaped by network activity. J Neurophysiol 93: 281–293

Lindsey BG, Morris KF, Shannon R, Gerstein GL (1997) Repeated patterns of distributed synchrony in neuronal assemblies. J Neurophysiol 78: 1714–1719

Louie K, Wilson MA (2001) Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. Neuron 29: 145–156

Maass W, Natschlager T, Markram H (2002) Real-time computing without stable states: a new framework for neural computation based on perturbations. Neural Comput 14: 2531–2560

Malamud BD, Morein G, Turcotte DL (1998) Forest fires: An example of self-organized critical behavior. Science 281: 1840–1842

Mao BQ, Hamzei-Sichani F, Aronov D, Froemke RC, Yuste R (2001) Dynamics of spontaneous activity in neocortical slices. Neuron 32: 883–898

Massimini M, Ferrarelli F, Huber R, Esser SK, Singh H, Tononi G (2005) Breakdown of cortical effective connectivity during sleep. Science 309: 2228–2232

Mechelli A, Price CJ, Friston KJ, Ishai A (2004) Where bottom-up meets top-down: neuronal interactions during perception and imagery. Cereb Cortex 14: 1256–1265

Nadasdy Z, Hirase H, Czurko A, Csicsvari J, Buzsaki G (1999) Replay and time compression of recurring spike sequences in the hippocampus. J Neurosci 19: 9497–9507

Natschlager T, Maass W (2005) Dynamics of information and emergent computation in generic neural microcircuit models. Neural Netw 18: 1301–1308

Netoff TI, Clewley R, Arno S, Keck T, White JA (2004) Epilepsy in small-world networks. J Neurosci 24: 8075–8083

Nicolis G, Prigogine I (1989) Exploring complexity : an introduction. New York: W.H. Freeman

Oram MW, Wiener MC, Lestienne R, Richmond BJ (1999) Stochastic nature of precisely timed spike patterns in visual system neuronal responses. J Neurophysiol 81: 3021–3033

Paczuski M, Maslov S, Bak P (1996) Avalanche dynamics in evolution, growth, and depinning models. Physical Review E Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics 53: 414–443

Rabinovich M, Volkovskii A, Lecanda P, Huerta R, Abarbanel HDI, Laurent G (2001) Dynamical encoding by networks of competing neuron groups: Winnerless competition. Physical Review Letters 8706

Ramón y Cajal S (1909) Histologie du système nerveux de l'homme & des vertébrés. Paris,: Maloine

Rolls ET (1990) Theoretical and neurophysiological analysis of the functions of the primate hippocampus in memory. Cold Spring Harb Symp Quant Biol 55: 995–1006

Rowe J, Friston K, Frackowiak R, Passingham R (2002) Attention to action: specific modulation of corticocortical interactions in humans. Neuroimage 17: 988–998

Schiff SJ, Jerger K, Duong DH, Chang T, Spano ML, Ditto WL (1994) Controlling chaos in the brain. Nature 370: 615–620

Segev R, Baruchi I, Hulata E, Ben-Jacob E (2004) Hidden neuronal correlations in cultured networks. Phys Rev Lett 92: 118102

Seung HS (1998) Continuous attractors and oculomotor control. Neural Networks 11: 1253–1258

Sinha S, Ditto WL (1999) Computing with distributed chaos. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics 60: 363–377

Skaggs WE, McNaughton BL (1992) Computational approaches to hippocampal function. Curr Opin Neurobiol 2: 209–211

Skaggs WE, McNaughton BL, Wilson MA, Barnes CA (1996) Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. Hippocampus 6: 149–172

Song S, Sjostrom PJ, Reigl M, Nelson S, Chklovskii DB (2005) Highly nonrandom features of synaptic connectivity in local cortical circuits. PLoS Biol 3: e68

Sporns O, Tononi G, Edelman GM (2000) Connectivity and complexity: the relationship between neuroanatomy and brain dynamics. Neural Netw 13: 909–922

Sporns O, Tononi G, Kotter R (2005) The human connectome: a structural description of the human brain. PLoS Comput Biol 1: e42

Sporns O, Zwi JD (2004) The small world of the cerebral cortex. Neuroinformatics 2: 145–162

Spors H, Grinvald A (2002) Spatio-temporal dynamics of odor representations in the mammalian olfactory bulb. Neuron 34: 301–315

Stam CJ, Jones BF, Nolte G, Breakspear M, Scheltens P (2006) Small-World Networks and Functional Connectivity in Alzheimer's Disease. Cereb Cortex

Stanley HE (1987) Introduction to phase transitions and critical phenomena. New York: Oxford University Press

Steinbuch K (1961) Die lernmatrix. Kybernetik 1: 36–45

Steinbuch K, Frank H (1961) [Non-digital learning matrices as perceptors.]. Kybernetik 1: 117–124

Stephan KE, Marshall JC, Friston KJ, Rowe JB, Ritzl A, Zilles K, Fink GR (2003) Lateralized cognitive processes and lateralized task control in the human brain. Science 301: 384–386

Strogatz SH (1994) Nonlinear dynamics and Chaos : with applications to physics, biology, chemistry, and engineering. Reading, Mass.: Addison-Wesley Pub

van Vreeswijk C, Sompolinsky H (1996) Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science 274: 1724–1726

Vogels TP, Rajan K, Abbott LF (2005) Neural network dynamics. Annu Rev Neurosci 28: 357–376

Wills TJ, Lever C, Cacucci F, Burgess N, O'Keefe J (2005) Attractor dynamics in the hippocampal representation of the local environment. Science 308: 873–876

Wilson MA (2002) Hippocampal memory formation, plasticity, and the role of sleep. Neurobiol Learn Mem 78: 565–569

Wilson MA, McNaughton BL (1993) Dynamics of the hippocampal ensemble code for space. Science 261: 1055–1058

Wolf A, Swift JB, Swinney HL, Vastano JA (1985) Determining Lyapunov Exponents from a Time-Series. Physica D 16: 285–317

Yeomans JM (1992) Statistical mechanics of phase transitions. Oxford [England] New York: Clarendon Press ; Oxford University Press

Structural Determinants of Functional Brain Dynamics

Olaf Sporns[1] and Giulio Tononi[2]

[1] Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405.
[2] Department of Psychiatry, University of Wisconsin, Madison, WI 53719.

1 Introduction

Neural connections of the mammalian cerebral cortex exhibit specific patterns ranging in scale from interconnections linking whole brain regions to intra-areal patterns of connections between cell populations or individual cortical neurons (Cajal, 1909; Brodmann, 1909; Zeki, 1993; Salin and Bullier, 1995; Swanson, 2003). Detailed anatomical and physiological studies have revealed many of the basic components and interconnections of cortical microcircuitry (Douglas and Martin, 1991), and of their arrangement into columns and mini-columns (Mountcastle, 1978; 1997). Columns and other localized populations of neurons maintain connections within and between brain regions, constituting large-scale patterns of anatomical connectivity. While the large-scale networks of human cortex remain largely unmapped (Sporns et al., 2005), comprehensive descriptions of anatomical patterns of cortical connectivity have been collated for several other mammalian species (e.g. Felleman and Van Essen, 1991; Scannell et al., 1999). Closer analysis has revealed that these patterns are neither completely regular nor completely random, but combine structural aspects of both of these extremes (reviewed in Sporns et al., 2004). This basic insight has sparked significant interest in characterizing the structure of brain networks, using methods that are also applied in parallel efforts to map and describe other biological networks, e.g. those of cellular metabolism, gene regulation, or ecology. This chapter is intended as an overview of recent quantitative approaches to brain networks (see also Sporns, 2005), with an emphasis on theoretical and computational studies that inform us about the structural features that determine functional brain dynamics.

In neuroscience, the term connectivity has multiple meanings and connotations that are sometimes difficult to define or disentangle (Horwitz, 2003; Lee et al., 2003). A fundamental distinction is that between structural, functional and effective connectivity, and we will adhere to this distinction for the remainder of the chapter. *Anatomical connectivity* is the set of physical or structural (synaptic) connections linking neurons within the network, as well

as their associated structural biophysical attributes encapsulated in parameters such as strength or effectiveness. Anatomical connections range in scale from local circuits to large-scale networks of inter-regional pathways. Their physical pattern may be thought of as relatively static at shorter time scales (seconds to minutes), but may be plastic or dynamic at longer time scales (hours to days), for example during learning or development. *Functional connectivity* (Friston, 1993; 1994) captures patterns of deviations from statistical independence between distributed and often spatially remote neuronal units, measuring their correlation/covariance, spectral coherence or phase-locking. Functional connectivity is highly time-dependent (on a scale of hundreds of milliseconds) and does not make any explicit reference to causal effects or an underlying structural model. *Effective connectivity* describes the network of causal effects of one neural system over another (Friston, 1994; Büchel and Friston, 2000), and can be inferred experimentally through perturbations or time series analysis. Unlike functional connectivity, effective connectivity is not "model-free", but usually requires the specification of a causal model including structural parameters.

The relationship between anatomical, functional and effective connectivity in the cortex currently represents one of the most significant challenges to computational cognitive neuroscience. An emerging view suggests that structural connection patterns are major determinants of the functional dynamics of cortical circuits and systems, as captured by functional or effective connectivity. According to this view, structural connections are essential for shaping patterns of activation and co-activation associated with specific cognitive states. Two potential linking principles are those of *segregation* and *integration* (Tononi et al., 1994; 1998; Friston, 2002; 2005). Segregation and integration are found throughout a broad range of cortical systems and may represent a set of complementary organizational principles. We will start this review by considering segregation and integration as basic principles, before turning to methods and approaches aimed at quantifying structural connection patterns, global measures of brain dynamics, and their interrelations.

2 Segregation and Integration

Anatomical and functional *segregation* refers to the existence of specialized neurons and brain areas, organized into distinct neuronal populations grouped together to form segregated cortical areas (Shipp and Zeki, 1985; Zeki, 1993). The concept of anatomical segregation is rooted in the notion that specific brain processes or functions can be localized to specific anatomical regions of the human brain, an idea that is central to the history of neurology and cognitive neuroscience (Phillips et al., 1984). Maps of cortical regions, such as those assembled by Ungerleider and Mishkin (1982), Van Essen and Maunsell (1983), Zeki and Shipp (1988), and Felleman and Van Essen (1991) have provided increasingly refined network diagrams of multiple anatomically

and functionally distinct areas of the primate visual cortex. These specialized and segregated brain regions contain neurons that selectively responded to specific input features (such as orientation, spatial frequency, or wavelength), or conjunctions of features (such as faces). Segregated areas maintain distinct patterns of connections with other areas, which are instrumental in defining these specialized local response properties (Passingham et al., 2002). Segregation can be found even within single cortical regions, where functionally distinct populations of neurons often remain spatially segregated. At least some intraregional (Gilbert and Wiesel, 1989; Tanigawa et al. 2005) and interregional (Angelucci et al., 2002) connections linking such populations are found to be patchy or clustered, preserving segregation.

Anatomical segregation entails that important correlates of specific functional brain states are found in localized changes of neuronal activity within specialized populations. However, segregated and specialized brain regions and neuronal populations must interact to generate functional dynamics. Coherent perceptual and cognitive states require the coordinated activation, i.e. the functional *integration*, of very large numbers of neurons within the distributed system of the cerebral cortex (Bressler, 1995; Friston, 2002). Electrophysiological studies have shown that perceptual or cognitive states are associated with specific and highly dynamic (short-lasting) patterns of temporal correlations (functional connectivity) between different regions of the thalamocortical system. Bressler has carried out numerous studies examining task-dependent large-scale networks of phase synchronization in primate and human cortex (Liang et al., 2000; Bressler and Kelso, 2001; Brovelli et al., 2004). Patterns of inter-regional cross-correlations have been found to accompany the performance of specific cognitive tasks in cats (e.g. Roelfsema et al., 1997), primates (Bressler, 1995) and humans (e.g. Srinivasan et al., 1999; Von Stein et al., 1999; Varela et al., 2001; Munk et al., 2002). McIntosh has documented changes in brain functional connectivity related to awareness (McIntosh et al., 1999), and most recently through recording differential interactivity of the human medial temporal lobe with other regions of the neocortex (McIntosh et al., 2003). Human neuroimaging experiments have revealed that virtually all perceptual or cognitive tasks, e.g. object recognition, memory encoding and retrieval, reading, working memory, attentional processing, motor planning and awareness are the result of activity within large-scale and distributed brain networks (McIntosh et al., 1999; 2000).

Common to most theoretical frameworks dealing with network aspects of cognition is the idea that integration across widely distributed brain regions requires neuronal interactions along inter-regional pathways. In the cortex, such interactions are mediated by the extensive and massive network of cortico-cortical connections. When these structural substrates of integration are disabled or disrupted, resulting in the disconnection of neuronal populations, specific functional deficits are often observed. While many observations suggest that disruptions of structural connections can result in deleterious effects on functional brain dynamics, we still lack a principled understanding

of how structural connections determine dynamics. In the brain, as in most other biological systems, structure and function are strongly interdependent – however, a comprehensive theoretical framework describing their interrelationship in the networks of the cerebral cortex remains elusive. In the remainder of this chapter, we focus on a set of measures that quantify structural connections and functional dynamics and we review several computational and empirical approaches that, utilizing such measures, aim at uncovering structural determinants of functional brain dynamics.

3 Measures of Structural Brain Connectivity

Neuronal networks consist of neurons connected by synapses. A major formal mathematical approach to the description of such networks is graph theory, especially the theory of directed graphs (Harary, 1969; Chartrand and Lesniak, 1996; Bang-Jensen and Gutin, 2001). Graphs have two major ingredients, nodes (cells, brain regions) and connections (synapses, pathways). In graph theory terminology, nodes are often referred to as vertices and connections as edges. Figure 1 provides an illustration of several elementary graph-theoretical concepts used in this chapter. In their simplest form, graphs can be described by a connection matrix or adjacency matrix with binary elements a_{ij} that represent the presence or absence of a directed edge between vertices j (source) and i (target). If such an edge exists, vertex j can directly communicate signals (spikes) to vertex i. It is important to note that in brain networks such direct connections are not the only way in which neuronal elements can influence each other. Indirect interactions can proceed along paths or cycles (Fig. 1A), defined as ordered sequences of distinct vertices and edges. More sophisticated and realistic formulations of networks as graphs include the weights or strengths of edges (Barrat et al., 2004a; 2000b). These quantitative approaches to weighted graphs have not yet been widely applied to neurobiological data sets.

The analysis of edges and paths within networks allow the quantification of a broad range of network characteristics, summarized in a series of recent reviews of complex networks (Strogatz, 2001; Albert and Barabási, 2002; Newman, 2003) and of brain networks (Hilgetag et al., 2002; Sporns, 2002; Sporns et al., 2004; Sporns, 2005). A wide spectrum of graph theory measures derive from the concepts of reachability and distance in graphs. The adjacency matrix of a network allows the derivation of the reachability matrix and the distance matrix, both fundamental for structural graph analyses. The reachability matrix indicates, for each ordered pair of vertices j and i, whether a path (of any length) exists from j to i. If all entries of the reachability matrix are ones, the network consists of only one component and is strongly connected. If the reachability matrix can be partitioned into non-overlapping subsets of vertices with no paths between them then the graph contains multiple (disconnected) components. Distance in a graph refers to the lengths

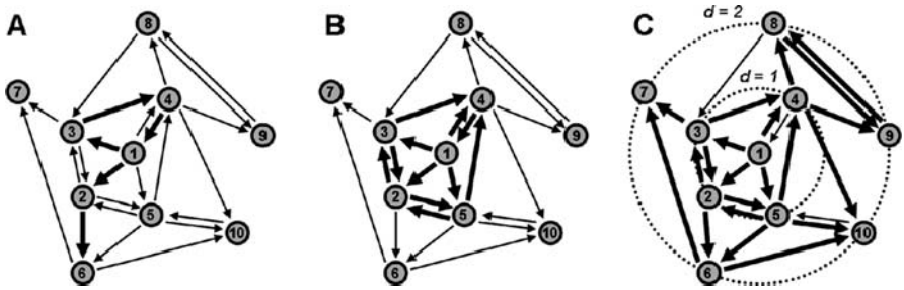


Fig. 1. Illustration of degrees, paths, cycles, clustering coefficient and hierarchical measures. All panels show an example of a digraph composed of N = 10 vertices and K = 23 directed edges. (**A**) Vertex 1 has indegree = 1, and outdegree = 4. Originating from vertex 1, bold edges indicate a path from vertex 1 to vertex 6, passing through vertex 2, with a length of 2, denoted as path {1,2,6}. The distance d_{61} is 2. Other valid paths from 1 to 6 include {1,5,6}, {1,3,2,6} and {1,4,9,8,3,2,5,6}. All paths consist of a set of unique vertices and edges, each visited only once. Another set of bold edges marks a cycle from vertex 1 to 3, 4 and back to 1, with a length of 3, denoted {1,3,4,1}. Other cycles are {1,4,1}and {1,3,2,5,4,1}. (**B**) Clustering coefficient of vertex 1. This vertex's neighbors are 2, 3, 4 and 5, and they maintain 6 connections among them out of 12 possible (4^2–4). This results in a clustering coefficient of 6/12=0.5. (**C**) Hierarchical clustering coefficient and divergence ratio for vertex 1 and $d = 1$, $d = 2$. Vertices 2, 3, 4 and 5 are at distance $d = 1$, and vertices 6, 7, 8, 9 and 10 are at distance $d = 2$ from vertex 1 (stippled circles). To compute the hierarchical measures, only outgoing connections linking successive hierarchical levels and intra-level connections are considered (bold arrows). Hierarchical clustering coefficients are 6/12 and 4/20 for levels d = 1 and d = 2, respectively. The divergence ratio D_1 is 5/7, given 7 outgoing connections between ring 1 and 2, and 5 vertices on ring 2. Modified after Sporns and Zwi (2004), and Da F. Costa and Sporns (2005)

of paths between vertices. The entries of the distance matrix give the length of the shortest (directed) path between the two vertices j and i. The global maximum of the distance matrix is also called the graph diameter.

Numerous measures of network connectivity can be derived from the adjacency matrix, the reachability matrix and the distance matrix of a graph. For example, the adjacency matrix allows the examination of the degree distribution of a given network. The degree of a vertex is the number of incident edges, sorted into indegree and outdegree, for incoming and outgoing edges, respectively (Fig. 1A). The degree distribution of a network provides important insights into whether the network contains vertices with approximately equal degrees (i.e. conforming to a Gaussian distribution) or whether the network's vertices show an exponential degree distribution. Such an exponential distribution is found when most of a network's vertices maintain few connections, while some of them are very heavily connected to large portions of the network (so-called hubs). Networks with Gaussian and exponential degree distributions are called "one-scale" and "scale-free", respectively (Amaral

et al., 2000), and may support very different dynamical behavior. Scale-free networks are found in many technological as well as biological systems, including metabolic and genetic regulatory networks. However, large-scale cortical networks (Sporns and Zwi, 2004) as well as networks of the brainstem reticular formation (Humphries et al., 2005) show little evidence of a scale-free architecture, perhaps due to the fact that major structural hubs cannot be easily accommodated given volume and metabolic limits. The functional interpretation of degree distributions for individual vertices is fairly straightforward. A high indegree indicates that the vertex is influenced by a large number of other vertices (a dynamical "sink"), while a high outdegree indicates a large number of potential functional targets (a dynamical "source"). The relation between the indegree and the outdegree of a vertex has been defined as the transmission index (Kötter and Stephan, 2003), expressed as the ratio between efferent edges (outdegree) and all known edges (indegree plus outdegree) of the vertex. In conjunction with other such vertex-specific indices, the transmission index allows a comparative analysis of the degree to which individual brain regions participate in network interactions. The examination of large-scale connectivity matrices indicates that each region's pattern of interactions is unique. This important fact was noted by Passingham et al. (2002) who suggested that this specific pattern may be crucial for determining the functional specificity of the region. The uniqueness of cortical regional connections led these authors to coin the term "connectional fingerprint" for the pattern of incident edges (connections) on brain regions.

While indegree and outdegree capture information about the local connectivity neighborhood of a given vertex, there are a number of measures that capture something about the global organization of a network. For example, small-world networks (Watts and Strogatz, 1998; Watts, 1999; 2003) combine features of regular and random graphs and appear to be ubiquitous within the natural, social and technological world (e.g. Strogatz, 2001, Albert and Barabási, 2002). The two main features of small-world networks are a high degree of local clustering and short average path lengths. Interestingly, these two features map onto two previously discussed candidate principles for cortical network organization, segregation and integration (Sporns et al., 2004). A high degree of local clustering in small-world networks is consistent with a high level of local segregation. The capacity to communicate between all their constituent vertices along short paths, measured as the characteristic path length, is consistent with global integration. The average (or median) of all the entries of the distance matrix constitutes the characteristic path length of a graph, $\lambda(G)$. The clustering coefficient of a vertex $\gamma(v)$ (Fig. 1B) indicates how many connections are maintained between this vertex's neighbors, defined as all those vertices that are connected to it, either through an incoming or an outgoing connection. The average of the clustering coefficients for each individual vertex is the clustering coefficient of the graph $\gamma(G)$.

These methods and measures for characterizing anatomical connection patterns have been applied to large-scale connectivity matrices of the cerebral

cortex, which have been assembled from hundreds of neuroanatomical studies conducted in a variety of species, including cat (Scannell et al., 1999) and nonhuman primates (Felleman and Van Essen, 1991; Young, 1993). Results indicate that the cerebral cortex is comprised of clusters of densely and reciprocally coupled cortical areas that are globally interconnected (Sporns et al., 2000a; 2000b). Regarding this clustered architecture, there is strong agreement between different clustering methods (Hilgetag et al., 2000; Sporns et al., 2000a). Importantly, large-scale cortical networks share some attributes of small-world networks, including high values for clustering coefficients and short characteristic path lengths (Hilgetag et al., 2000; Sporns et al., 2000a). A recent detailed analysis revealed that these properties are shared by large-scale connection matrices from several species and cortical systems, and are also found in connection matrices generated by making empirically based probabilistic assumptions about local connection densities and arborizations of cortico-cortical connections (Sporns and Zwi, 2004). Interestingly, self-similar or fractal connection matrices also exhibit small-world connectivity patterns, in addition to a number of other characteristics in line with known features of cortical connectivity, giving rise to the hypothesis that self-similar connectivity may also be found in cortex (Sporns, 2006).

Degree and clustering coefficient evaluate characteristics of connectivity within the immediate topological neighborhood a vertex. A natural extension of such measures involves their application to hierarchical levels around a given vertex (Fig. 1C), defined as the set of vertices that can be reached by minimum paths of increasing lengths d (Da F. Costa, 2004). This generalization allows the calculation of hierarchical measures that capture a much broader context around each vertex, thus more accurately defining how the vertex is embedded in and contributes to the overall network. For example, the divergence ratio D_d of a central vertex is defined as the ratio between the number of hierarchical neighbors at a given distance $d+1$ and the hierarchical degree (the number of connections linking vertices at distances d and $d+1$). Similarly, the clustering coefficient can be generalized to apply to a given hierarchical level, with the definition given in the previous paragraph representing the limit case of $d = 1$. Application of such hierarchical measures to large-scale cortical networks has revealed statistically significant differences between groups of brain regions, such as the dorsal and ventral visual pathway in the macaque (Da F. Costa and Sporns, 2005).

While the existence of paths in brain networks allows for potential interactions across longer distances, a realistic assumption is that much of the processing characteristics and functional contributions of a vertex is determined by its interactions within a local neighborhood (defined in terms of topological, not necessarily metric, distance). To aid in the analysis of such local neighborhoods, large networks or graphs can be decomposed into smaller "building blocks" or "networks-within-networks". Such subgraphs or motifs (Milo et al., 2002; 2004) form a basic structural alphabet – for example, given three vertices, there are only 13 distinct ways to interconnect them (Fig. 2). Motifs

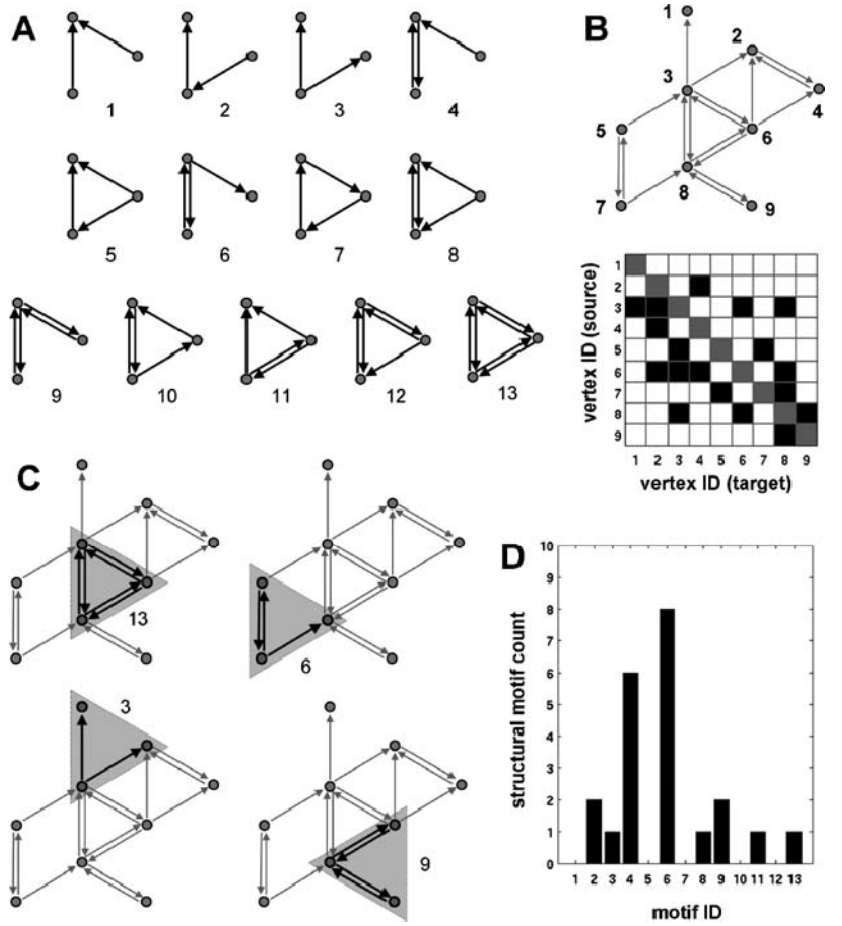


Fig. 2. Definition and detection of structural motifs. (**A**) Complete set of 13 motifs for 3 vertices. Number refers to the motif class (motif ID). (**B**) Example digraph with 9 numbered vertices and 18 edges. Graph (top) and adjacency matrix (bottom) are shown. Entries on the main diagonal of the adjacency matrix are shown in gray, edges are shown in black. (**C**) Motif detection within graph shown in panel B. Instances of motif classes 13, 6, 3 and 9 are highlighted. (**D**) Complete motif frequency spectrum for the graph shown in panel B

occur in characteristic numbers and distributions that can be highly characteristic of and informative about the large-scale structural and functional characteristics of the global network. Milo et al., reported that some specific motifs are statistically increased in biological networks, as compared to equivalent random networks. Large-scale cortical networks also contain specific motifs in greater than expected abundance, shared across at least two mammalian species (Sporns and Kötter, 2004; Fig. 3). This analysis also revealed that, globally, large-scale brain networks contain relatively few structural motifs

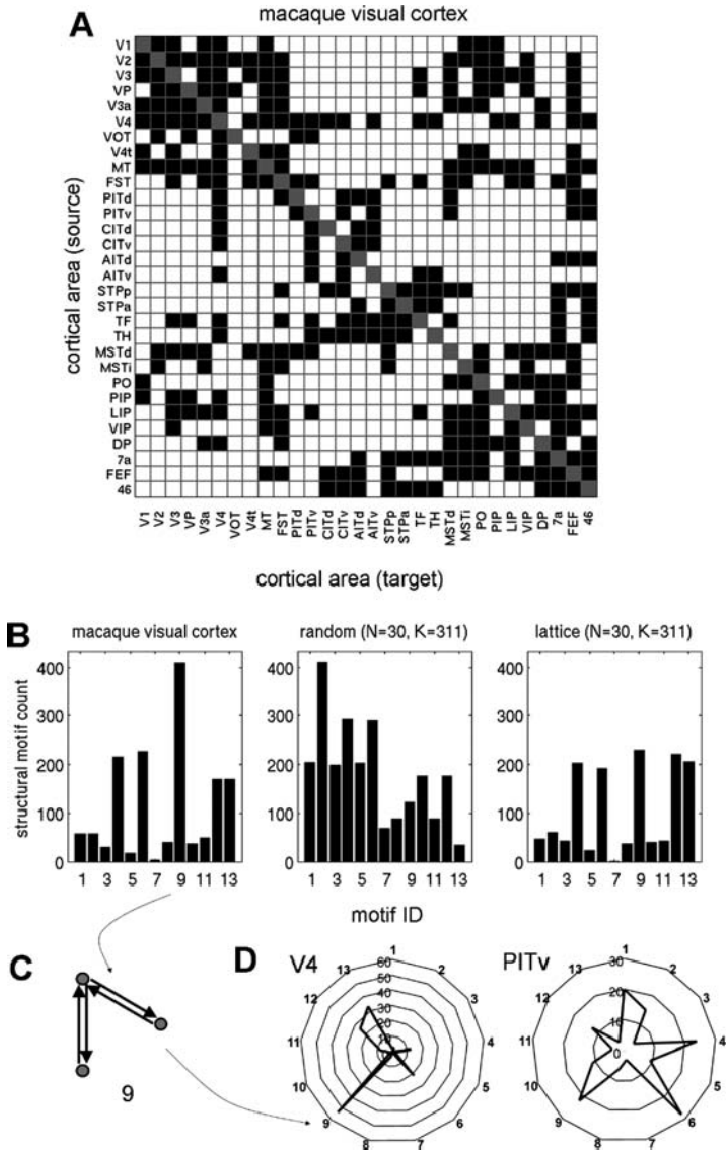


Fig. 3. Structural motifs in macaque visual cortex. (**A**) Adjacency matrix of macaque visual cortex. (**B**) Structural motif frequency spectrum for macaque visual cortex (left), as well as equivalent random matrices (middle) and equivalent lattice matrices (right). (**C**) Motif 9 is the only structural motifs that is significantly increased over both random and lattice networks. (**D**) Motif fingerprints of areas V4 and PIVv demonstrate that individual brain regions make different contributions to the overall motif frequency spectrum shown in panel B. V4 is one of only a few areas that show very high proportions of motif 9. Modified after Sporns and Zwi (2004), and Sporns and Kötter (2004)

compared to randomized controls, while at the same time maximizing the number of potential functional patterns. More recently, motif analysis has also been applied to single neuron networks between layer 5 cortical pyramidal neurons (Song et al., 2005), confirming the non-randomness of neural connections at this organizational level. A significant further extension of the concept of motifs was introduced by Onnela et al. (2005), who derived measures of motif intensity and coherence which allow motif counts to be applied to weighted networks.

An important issue concerns the connection between motifs and the kind of dynamics they support. Zhigulin (2004) developed an approach to extracting and quantifying dynamical motifs, defined as small subnetworks with nontrivial dynamics. Zhigulin observed that the appearance of specific types of dynamics in large networks was linked to the appearance of periodic and chaotic dynamical motifs. Prill et al. (2005) have drawn relationships between motif patterns and a specific dynamical property, stability or robustness to small perturbations. These authors argue that at least some of the non-randomness of biological networks can be explained by adaptive advantages conferred by robust dynamical stability.

A large literature centers on how complex natural or technological networks are affected by damage to their connection pattern. The world wide web, an example of a scale-free network (Barabási and Albert, 1999), has been shown to be surprisingly robust with respect to random deletion of nodes, but rather vulnerable to targeted attack on heavily connected hubs (Albert et al., 2000; Doyle et al., 2005), which often results in disintegration of the network. In the brain, the mapping of functional deficits to underlying structural perturbations is experimentally challenging, but essential for a more complete understanding of brain damage and recovery. It is currently unknown which structural measures best capture the potential effects of vertex or edge lesions, although candidate measures of edge vulnerability (Kaiser and Hilgetag Kaiser, 2004) have been defined and have led to the identification of edges whose loss most affects global structural measures. Such edges often correspond to "bridges", i.e. edges linking segregated clusters of brain regions. The issue of defining measures of robustness or vulnerability in brain networks is conceptually linked to the problem of objectively defining the functional contributions of individual network elements (Keinan et al., 2004).

Finally, we should note that measures of structural, functional and effective connectivity increasingly intersect, as in the analysis of functional or effective connectivity patterns as graphs (Dodel et al., 2002; Salvador et al., 2005a; Eichler, 2005). Applications of connectivity analyses to EEG, MEG and fMRI data sets have been reviewed in several other chapters in this volume (Feree and Nunez, 2007; Darvas and Leahy, 2007; Bressler and McIntosh, 2007). Essentially, patterns of cross-correlation or coherence can be conceptualized as undirected graphs with edges that represent the existence and, in some cases, the strength of the statistical relationship between the linked vertices. Studies of patterns of functional connectivity (based on coherence or

correlation) among cortical regions have demonstrated that functional brain networks exhibit small-world (Stam, 2004; Salvador et al., 2005b; Achard et al., 2006; Salvador et al., 2007) and scale-free properties (Eguiluz et al., 2005), possibly reflecting the underlying structural organization of anatomical connections. For example, it is an open question if nodes in structural and functional neuronal connectivity matrices maintain similar patterns of connectivity and exhibit similar local properties such as clustering.

4 Measures of Brain Dynamics: Functional Connectivity

As many of the structural studies reviewed in the previous section illustrate, brain networks (like other biological networks) are neither completely random nor completely regular. Instead their local and global structure exhibits significant departures from randomness. A key question concerns how these nonrandom features of brain structural connectivity relate to brain function or dynamics. A consideration of brain evolution may guide our answer. In the course of evolution, brain connectivity is one of the prime substrates, the gradual modification of which in an adaptive context contributes to enhanced fitness and survival. Biological structure/function relationship often become more comprehensible when viewed in the context of evolution, for example when we consider the structure and function of proteins, cellular organelles, or entire body plans. The evolutionary history of the primate and especially human brain may ultimately hold the key for understanding the structural basis of cognition (for a modern review of brain evolution, see Striedter, 2005).

As we approach the question of how structure determines function in the brain, we turn next to measures of brain dynamics based on functional connectivity. As outlined in other chapters of this volume (e.g. Jirsa and Breakspear, this volume) there are numerous approaches to quantifying or measuring brain dynamics. In this chapter, we will focus on measures that attempt to capture global aspects of functional connectivity, i.e. patterns of statistical dependence between often remote neural units or brain regions (Friston, 1993), building on the firm foundation offered by statistical information theory (Cover and Thomas, 1991; Papoulis, 1991). In its most general form, statistical dependence is expressed as an estimate of mutual information. Unlike correlation, which is a linear measure of association, mutual information captures all linear or nonlinear relationships between variables. While the mathematical definition of mutual information is quite straightforward, the actual derivation of valid estimates for entropy and mutual information for any given application can be challenging and is the subject of much ongoing research.

Mutual information between two units A and B is defined as the difference between the sum of their individual entropies and their joint entropy:

$$\mathrm{MI(A,B)} = \mathrm{H(A)} + \mathrm{H(B)} - \mathrm{H(AB)} \tag{4.1}$$

Note that $MI(A,B) \geq 0$ and $MI(A,B) = MI(B,A)$. The mutual information MI(A,B) will be zero if no statistical relationship exists between A and B, i.e. if A and B behave statistically independently such that $H(AB) = H(A) + H(B)$. The upper bound for mutual information between A and B is given by the lesser of the two entropies. Mutual information expresses the amount of information that the observation of one unit conveys about the other unit. Any reduction of the joint entropy H(AB) such that $H(AB) < H(A) + H(B)$ indicates some degree of statistical dependence between A and B and will result in a positive value for the mutual information.

Mutual information has certain limitations. First, we note that the existence of positive mutual information between A and B does not express causal influences from A on B or *vice versa*. Hence, mutual information is informative in the context of functional connectivity, but does not allow (by itself) the inference of effective connectivity. Furthermore, in any real or simulated system, the estimation of mutual information critically depends on correct estimates for the individual and joint entropies, which in turn are often derived from their respective state probability distributions. As mentioned above, these estimates can be difficult to derive from small or sparse data sets such as those often encountered in neurobiological applications and, in many cases, additional statistical assumptions have to be made (e.g. Paninski, 2003; Pola et al., 2003).

In (4.1) A and B refer to individual variables representing individual neurons or brain regions. Mutual information can also be defined within larger systems. For example, let us consider a system X composed of n elements that is partitioned into two complementary subsets of elements. One subset consists of k elements and is denoted as X^k, while its complement contains the remaining $n-k$ elements and is denoted as $X - X^k$. The mutual information between these two subsets is

$$MI(X^k, X - X^k) = H(X^k) + H(X - X^k) - H(X) \tag{4.2}$$

While mutual information captures the degree of statistical dependence between two elements (or subsets), the integration I(X) of a system measures the total amount of statistical dependence among an arbitrarily large set of elements (Tononi et al., 1994). As the definition (4.3) illustrates, integration can be viewed as the multivariate generalization of mutual information. Considering a system X, composed of a set of elements $\{x_i\}$, its integration I(X) is then defined as the difference between the sum of the entropies of the individual elements and their joint entropy:

$$I(X) = \sum_i H(x_i) - H(X) \tag{4.3}$$

Given this definition, integration essentially quantifies the divergence between the joint probability distribution of the system X and the product of the marginal distributions of the individual elements (Schneidman et al., 2003a;

McGill, 1954). This measure has also been called the multi-information, as it expresses the total information shared by *at least two or more* elements. Integration (multi-information) differs from another multivariate informational measure called the co-information (Bell, 2003), which captures only the information shared by *all* elements. Venn diagrams illustrating the relationship between mutual information, integration and co-information are shown in Fig. 4. Similar to mutual information, integration may be viewed as the amount of error one makes given the assumption of independence between all variables. Note further that, like mutual information, $I(X) \geq 0$. If all elements are statistically independent their joint entropy is exactly equal to the sum of the

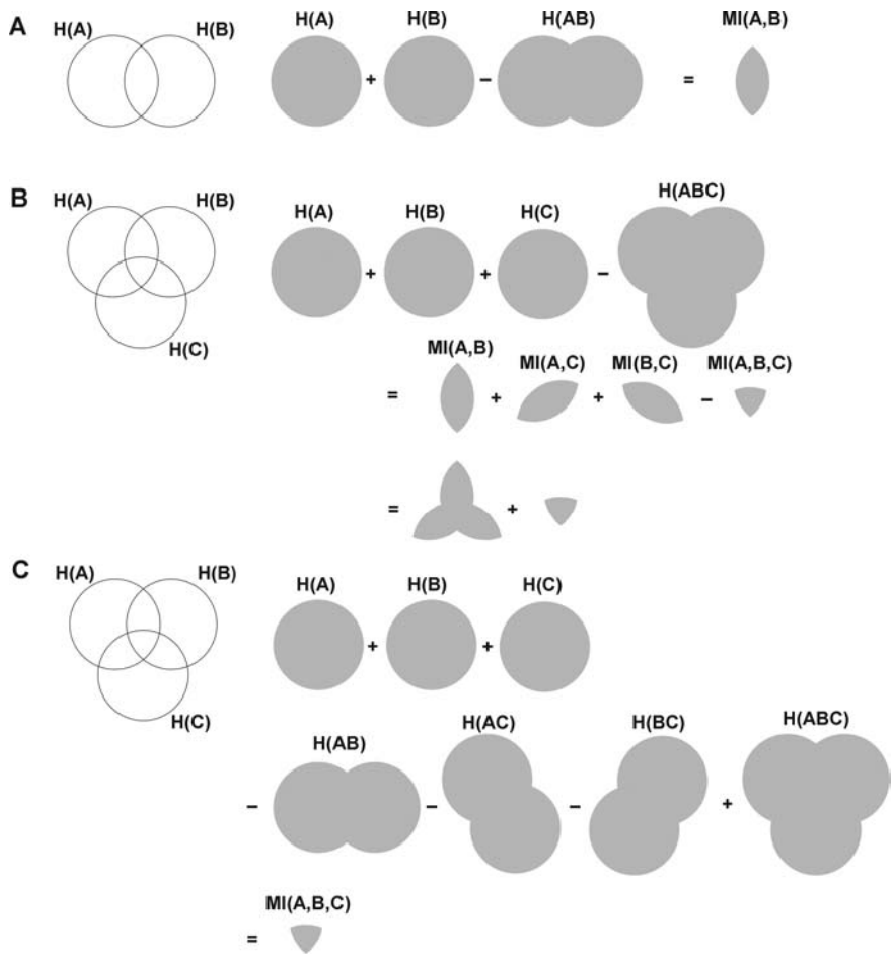


Fig. 4. Venn diagrams illustrating the relationship between mutual information (**A**), integration or multi-information (**B**) and co-information (**C**). Integration and co-information are shown for systems of 3 units, but readily generalize to larger systems

element's individual entropies and $I(X) = 0$. Any amount of statistical dependence between the elements will express itself in a reduction of the element's joint entropy and thus in a positive value for $I(X)$. As is the case for mutual information, an upper bound for $I(X)$ can be calculated from the spectrum of the individual entropies. In summary, integration quantifies the total amount of statistical structure or statistical dependencies present within the system.

Given a system of size n, we can define integration not only for the entire system but also for all hierarchical levels $k < n$ within it. We denote the average integration of subsets of size k as the hierarchical integration $< I(X^k) >$, noting that under random sampling the average is taken over all k-out-of-n subsets. Thus, $< I(X^n) >= I(X)$ and $< I(X^1) >= 0$. It can be proven that for any given system the spectrum of average integration for all values of k ($1 \leq k \leq n$) must increase monotonically, i.e. $< I(X^{k+1}) > \geq < I(X^k) >$. The difference between successive levels $< I(X^{k+1}) > - < I(X^k) >$ increases and approaches a constant value, indicating that the amount of integration (statistical dependence) gained by adding further elements to the system approaches a limit. Intuitively, this characteristic of hierarchical integration reflects similar properties described for informational measures of population redundancy (Schneidman et al., 2003b; Puchalla et al., 2005).

The characterization of the spectrum of average integration across all levels of scale (subset size k) within a given system allows us to examine how and where statistical structure within the system is distributed. How is this possible? Let us say that we find that a system as a whole has a certain amount of statistical structure, measured by its integration $I(X) > 0$. This means that some statistical dependencies exist somewhere, at some spatial scale, within the system X. But the global estimate of $I(X)$ does not provide information as to whether this structure is homogeneously distributed throughout the system, or whether this structure is localized or "concentrated" among specific units or subsets. If statistical dependencies are homogenously distributed, the system would be, in terms of its functional connectivity, totally undifferentiated, essentially presenting the same view to an observer zooming in on different levels of scale. We might say that such as system lacks any functional segregation. If statistical dependencies exist predominantly within subsets of specific size, there would be parts of the system that are more integrated than others and these integrated subsets would represent local structure. Such a system contains functional segregation in addition to the global functional integration expressed by $I(X)$ at the highest level.

To differentiate between these possibilities, we need a measure that takes into account the full distribution of integration across levels of scale (Fig. 5). Such a measure, which captures the extent to which a system is both functionally segregated (small subsets of the system tend to behave independently) and functionally integrated (large subsets tend to behave coherently), was proposed by Tononi et al. (1994). This statistical measure, called neural complexity $C_N(X)$, takes into account the full spectrum of subsets and can be derived either from the ensemble average of integration for all subset sizes 1

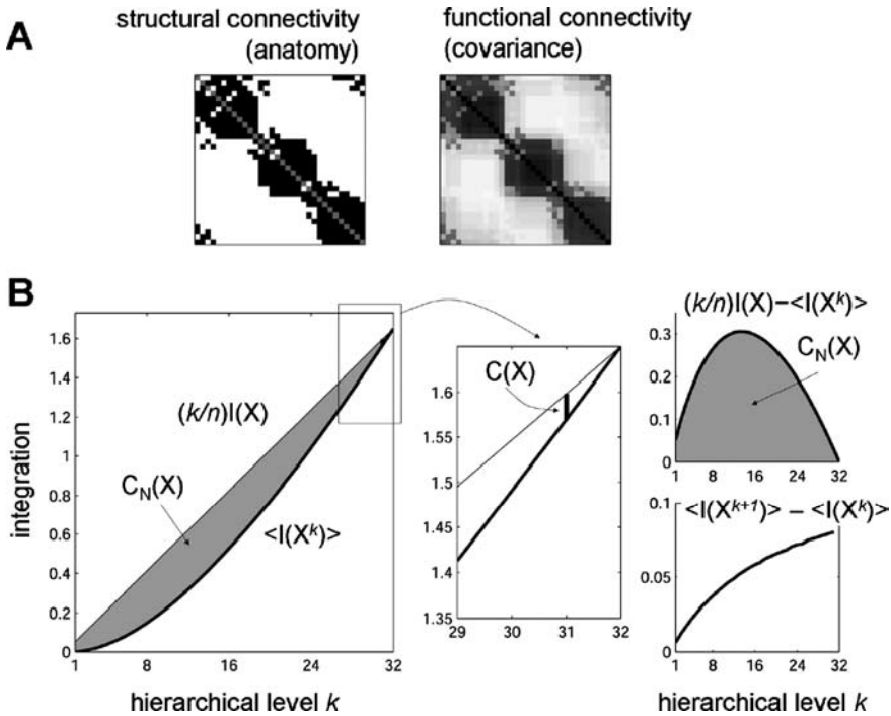


Fig. 5. Complexity $C_N(X)$ and C(X) for an example network of 32 units, generated by optimizing C(X) as described in Sporns et al., 2000a. (**A**) Structural and functional connectivity pattern. (**B**) Full spectrum of hierarchical integration (levels 1 to n), with neural complexity $C_N(X)$ corresponding to the shaded area. Inset at right shows a magnified part of the spectrum around levels 29 to 32, with C(X) corresponding to the difference between hierarchical integration profiles at level $n-1$. Rightmost plots show an alternative way of plotting complexity emphasizing a maximal difference in hierarchical integration profiles at a specific level (top), and the accelerating increase in hierarchical integration between successive levels (bottom)

to n, or (equivalently) from the ensemble average of the mutual information between subsets of a given size (ranging from 1 to $n/2$) and their complement. $C_N(X)$ is defined as:

$$\begin{aligned} C_N(X) &= \Sigma_k (k/n) I(X) - < I(X^k) > \\ &= \Sigma_k < MI(X^k, X - X^k) > \end{aligned} \tag{4.4}$$

As is evident from the second expression for $C_N(X)$, the complexity of a system is high when, on average, the mutual information between any subset of the system and its complement is high. The hierarchical nature of this measure of complexity spanning all levels of scale within the system is inherently well suited for a system such as the brain, which is characterized by modularity at several different levels, ranging from single neurons to brain regions. Thus, complexity is complementary to recent approaches that investigate brain

dynamics in the context of a nested multilevel, multiscale architecture (Breakspear and Stam, 2005).

Another closely related but nonidentical measure of complexity expresses the portion of the entropy that is accounted for by the interactions among all the components of a system (Tononi et al., 1998; Tononi et al., 1999; Fig. 5). There are three mathematically equivalent expressions for this measure, called C(X):

$$\begin{aligned} \mathrm{C(X)} &= \mathrm{H(X)} - \Sigma_i \mathrm{H}(\mathrm{x}_i \big| \mathrm{X} - \mathrm{x}_i) \\ &= \Sigma_i \mathrm{MI}(\mathrm{x}_i, \mathrm{X} - \mathrm{x}_i) - \mathrm{I(X)} \\ &= (n-1)\mathrm{I(X)} - n < \mathrm{I}(\mathrm{X} - \mathrm{x}_i) > \end{aligned} \tag{4.5}$$

$\mathrm{H}(\mathrm{x}_i \big| \mathrm{X} - \mathrm{x}_i)$ denotes the conditional entropy of each element x_i, given the entropy of the rest of the system $\mathrm{X} - \mathrm{x}_i$. We note that $\mathrm{C_N(X)}$ as well as C(X) are always greater or equal to zero. Both $\mathrm{C_N(X)}$ as well as C(X) will be exactly zero for systems with zero integration (no statistical dependence at any level), and they will be small (but non-zero) for systems that have non-zero integration, but for which this integration is homogeneously distributed within the system.

While the third formulation of C(X) has a straightforward graphical interpretation (Fig. 5), the second formulation of C(X) is perhaps most useful to provide an intuitive computational basis for this measure. C(X) is obtained as the difference of two terms: the sum of the mutual information between each individual element and the rest of the system minus the total amount of integration. Thus, C(X) takes on large values if single elements are highly informative about the system to which they belong, while not being overly alike (as they would tend to be if their total integration, or total shared information, is high). $\mathrm{C_N(X)}$ and C(X) are closely related (Fig. 5B), but not mathematically equivalent.

Within the context of applications of brain functional connectivity, it is essential to underscore that complexity captures the degree to which a neural system combines functional segregation and functional integration. Extensive computational explorations (Tononi et al., 1994; 1998; Sporns et al., 2000a; 2000b; Sporns and Tononi, 2002) have shown that complexity is high for systems that contain specialized elements capable of global (system-wide) interactions. On the other hand, complexity is low for random systems, or for systems that are highly uniform, corresponding to systems that lack either global integration or local specialization, respectively. The relation of connectivity topology and complexity has recently been analytically investigated (De Lucia et al., 2005).

5 Measures of Brain Dynamics: Effective Connectivity

Measures based on mutual information are useful for analyzing functional connectivity patterns, obtained from neuronal spike trains, local field potential recordings or fMRI/PET voxel time series. However, functional connectivity

allows only very limited insights into patterns of causal interactions within the network. Patterns of functional connectivity are statistical signatures of hidden causal processes occurring within and among specific and time-varying subsets of neurons and brain regions. The identification of which subsets are currently causally engaged in a given task requires the inclusion of and reference to a structural model in order to access effective connectivity patterns.

Effective connectivity attempts to reconstruct or "explain" recorded time-varying activity patterns in terms of underlying causal influences of one brain region over another (Friston, 1994; Büchel and Friston, 2000; Lee et al., 2003). This involves the combination of (essentially covariance-based) functional connectivity patterns with a structural system-level model of interconnectivity. A technique called "covariance structural equation modeling" is used to assign effective connection strengths to anatomical pathways that best match observed covariances in a given task (McIntosh and Gonzalez-Lima, 1994; Horwitz et al., 1999). Applied in different cognitive tasks, this technique allows the identification of significant differences in effective connectivity between a given set of brain regions, illustrating the time- and task-dependent nature of these patterns. Another approach called "dynamic causal modeling" (Friston et al., 2003; Stephan and Friston, 2007) uses a Bayesian framework to estimate and make inferences about interregional influences, explicitly in the context of experimental changes. A caveat concerning these and other approaches to extracting effective connectivity is that they usually require assumptions about the identity of participating brain regions and the patterns and direction of cross-regional influences between them.

Another approach to identifying highly interactive brain regions and their causal interactions involves the use of effective information, a novel measure of the degree to which two brain regions or systems causally influence each other (Tononi, 2001; Tononi and Sporns, 2003). Given a neural system that is partitioned into two complementary subsets, A and B, we obtain the effective information from A to B by imposing maximal entropy on all outputs of A. Under these conditions the amount of entropy that is shared between A and B must be due to causal effects of A on B, mediated by connections linking A and B. These connections can either be direct connections crossing the bipartition or indirect links via a surrounding neural context. The effective information from A to B may then be formulated as

$$\mathrm{EI}(\mathrm{A} \rightarrow \mathrm{B}) = \mathrm{MI}(\mathrm{A}^{\mathrm{Hmax}}, \mathrm{B}) \tag{5.1}$$

Note that unlike MI(A,B), effective information may be non-symmetrical, i.e. $\mathrm{EI}(\mathrm{A} \rightarrow \mathrm{B}) \neq \mathrm{EI}(\mathrm{A} \rightarrow \mathrm{B})$, owing to non-symmetrical connection patterns. Furthermore, the estimation of effective information requires perturbations of units or connections.

It has been suggested that the integration of information is essential for the functioning of large-scale brain networks (e.g. Tononi et al., 1998; Tononi and Edelman, 1998). In considering information integration the notion of causality, or effectiveness, is crucial. A system that integrates information effectively

must do so via actual causal interactions occurring within it. Mere statistical coincidences are insufficient to characterize the participating entities as truly integrated. Tononi and Sporns (2003) developed a measure for information integration (called Φ) based on effective information that captures the maximal amount of information that can be integrated within the system. For a given system or system subset S composed of subsets A and B, Φ is defined as the capacity for information integration, or $\Phi(S)$, given by the value of $EI(A \leftrightarrow B)$ for the minimum information bipartition (MIB):

$$\Phi(S) = EI(MIB(S)) \tag{5.2}$$

This measure allows the simultaneous quantification of information integration as well as the identification of all those system elements that participate in it. It can thus be used to delineate integrated functional clusters or networks of effective connectivity from among larger sets of brain regions. It is important to note that, following this definition, information integration takes place within complexes, defined as subsets of elements capable of integrating information that are not part of any subset having higher Φ.

Currently, this measure of information integration has only been tested in computer simulations of small model systems with varying anatomical architectures (Tononi and Sporns, 2003; Fig. 6). The results indicate that information integration is maximized by two main attributes of the anatomical connection pattern. First, each element maintains a different connection pattern, or connectional "finger-print", a property that strongly promotes regional functional specialization. Second, the pattern maintains global connectedness and ensures that a large amount of information can be exchanged across any bipartition of the network, which in turn promotes global functional integration. Simple models of the connectional organization of specific neural architectures, such as the thalamocortical system, are found to be well suited to information integration, while others, such as the cerebellum, are not. Neural architectures that are highly capable of integrating information are also associated with consciousness. Tononi (2004) has suggested that consciousness critically depends on the ability of a neural substrate to integrate information and is therefore tied to specific and quantifiable aspects of effective brain connectivity.

Several other methods for analyzing causal influences in the brain have been proposed, many of which utilize the temporal dynamics of the observed neural system to extract information about effective interactions, building on the fundamental fact that causes must precede effects in time (for a comparative computational study see Lungarella et al., 2007). Several methods are based on interpretations or adaptations of the concept of Granger causality (Granger, 1969), involving estimates of how much information a set of variables provides for the prediction of another. For example, Kaminski et al. (2001) develop an approach based on exploiting the directed transfer function between two neural signals. Granger causality has been applied to EEG data sets obtained from large-scale sensorimotor networks (Brovelli et al., 2004) as

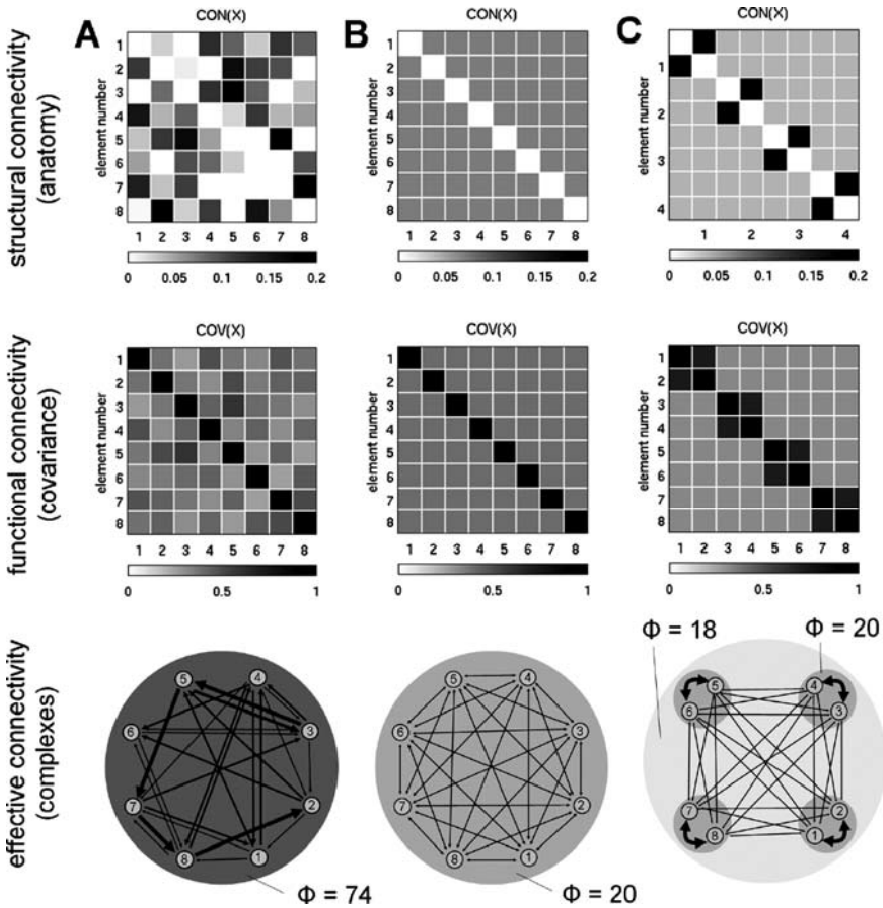


Fig. 6. Information integration. All panels show structural connectivity (top), functional connectivity (middle) and effective connectivity (bottom) for networks of 8 units. (**A**) Network obtained after optimizing for Φ, resulting in a single complex with $\Phi = 74$ bits. Structural connections are heterogeneous and span the entire network, jointly maximizing functional segregation and functional integration. (**B**) Uniform network (loss of functional segregation), with greatly reduced $\Phi = 20$ bits. (**C**) Modular network (loss of functional integration), split into four identical complexes with $\Phi = 20$ bits each. Modified after Tononi and Sporns (2003), and Tononi (2004)

well as fMRI time series (Roebroeck et al., 2005). Additional causality measures can discriminate between direct causality and effects mediated through extraneous system components (see also Liang et al., 2000). Bernasconi and König (1999) developed statistical measures that allowed the detection of directed dependences within temporal brain data sets. Schreiber (2000) defined a measure called transfer entropy, which is able to detect directed exchange of information between two systems by considering the effects of the state of

one element on the state transition probabilities of the other element. This yields a non-symmetric measure of the effects of one element on the other, exploiting the entire system's temporal dynamics.

The experimental application of measures of effective connectivity presents a number of difficult problems. Structural equation modeling and dynamic causal modeling are sensitive to choices made about the underlying structural model, while causal measures based time series analysis are prone to issues surrounding sample sizes or systematic sampling biases. Effective information, as defined above, shares some of these problems, in addition to issues related to its use of systematic perturbations, which are likely to be difficult to estimate in real neuronal systems. These difficulties notwithstanding, some promising avenues towards extracting effective connectivity from brain data have recently been pursued. The combination of transcranial magnetic stimulation (TMS) with functional neuroimaging, for the first time, allows the quantification of effects of localized perturbations on extended brain networks engaged in the performance of specific tasks (Paus, 1999; Pascual-Leone et al., 2000). Using a combination of TMS and high-density electroencephalography Massimi et al. (2005) reported a striking reduction in the extent of cortical effective connectivity during non-REM sleep compared to waking. This state-dependent difference is recorded in the same individual, presumably existing within an identical structural connectivity pattern. A major implication of this breakdown of effective connectivity during non-REM sleep is that it points towards a crucial role of causal influences between brain regions associated with information integration as a neural basis for consciousness (Tononi, 2004).

6 Relating Connectivity and Dynamics

How do the different dimensions of brain connectivity relate to one another? Answering this question demands the combined manipulation and analysis of structural and dynamic attributes. In this final section of the chapter, we briefly review several recent lines of research that have attempted to bridge structural, functional and effective connectivity with the use of computational modeling.

A crucial first step in linking structure to function involves the identification of functionally integrated and structurally connected networks that are potential building blocks of cognitive architectures. Effective integration of neural activity requires causal interactions, which must operate through physical connections. In fact, structural connectivity provides a first approach towards determining potential functional units, by revealing connectedness and modularity within graphs, as defined above. A next step is the identification of functional clusters, using informational measures (Tononi et al., 1998b; Sporns et al., 2000a) or through the application of standard clustering techniques to functional connectivity patterns. Finally, effective connectivity measures such as information integration (Tononi and Sporns, 2003) can aid

in the delineation of causally linked neural clusters and complexes. Mapping such clusters in the course of cognitive function would help to identify brain regions that are generating specific cognitive states and discriminate them from others that are activated, but not causally engaged.

Computational approaches allow the systematic study of how neural dynamics is shaped by the structure of connection patterns linking individual elements (Fig. 7). This has been investigated in detailed computer simulations of cortical networks with heterogeneous (Jirsa and Kelso, 2000; Jirsa 2004; Assisi et al., 2005) and spatially patterned (Sporns, 2004) connection topologies. It was found that different connection topologies generated different modes of neuronal dynamics, and some systematic tendencies could be identified. For example, connectivity patterns containing locally clustered connections with a small admixture of long-range connections were shown to exhibit robust small-world attributes (Sporns and Zwi, 2004; Sporns, 2004; Kaiser and Hilgetag, 2006), while conserving wiring length. These connectivity patterns also gave rise to functional connectivity of high complexity with heterogeneous spatially and temporally highly organized patterns. These computational studies suggest the hypothesis that only specific classes of connectivity patterns (which turn out to be structurally similar to cortical networks) simultaneously support short wiring, small-world attributes, clustered architectures (all structural features), and high complexity (a global property of functional connectivity).

The discovery of small-world connectivity patterns in functional connectivity patterns derived from fMRI, EEG and MEG studies (Stam, 2004; Salvador et al., 2005b; Achard et al., 2006; Bassett and Bullmore 2006; Salvador et al., 2007) raises the question of how closely functional connections map onto structural connections. The state- and task-dependence of functional connectivity suggests that a one-to-one mapping of structural to functional connections does not exist. However, it is likely that at least some structural characteristics of individual nodes are reflected in their functional interactions – for example, hub regions should maintain larger numbers of functional relations. A variety of neuro-computational models have suggested that small-world connectivity imposes specific constraints on neural dynamics at the large scale. Numerous studies suggest that small-world attributes facilitate synchronization and sustained activity, irrespective of the details of the node dynamics that are employed in the model (Nishikawa et al., 2003; Buszaki et al., 2004; Masuda and Aihara, 2004; Netoff et al., 2004; Roxin et al., 2004). Synchronization-based rewiring rules promote the emergence of small-world architectures from random topologies (Gong and van Leeuwen, 2004), underscoring the reciprocal "symbiotic" relationship between neural dynamics and underlying brain architectures (Breakspear et al., 2006). Plasticity rules shape structural connectivity, resulting in neural dynamics that in turn shapes plasticity.

Yet another interesting connection between structural connectivity and global dynamics is based on the idea that the continual integration and redistribution of neuronal impulses represents a critical branching process (Beggs

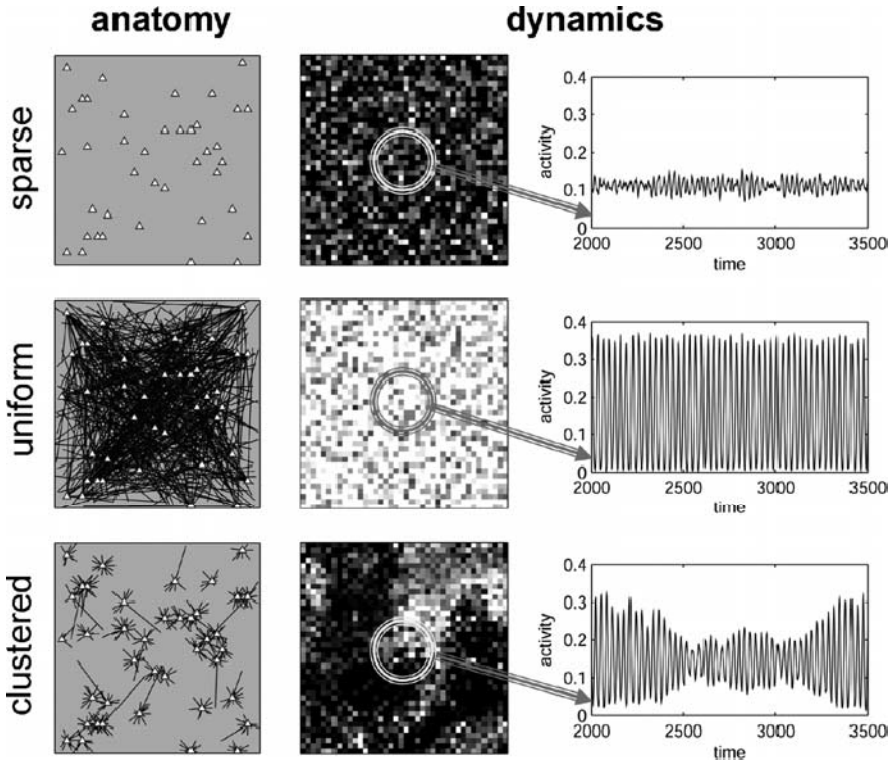


Fig. 7. Relation of structural connectivity and functional dynamics in intermediate-scale cortical networks. The model consists of a single map of 40 × 40 nonlinear neuronal units, each modeled as a single reciprocally coupled Wilson-Cowan excitatory and inhibitory unit, interconnected in different patterns. Three patterns of structural connectivity are shown: "sparse" (intra-map connections absent), "uniform" (intra-map connections are assigned at random with uniform probability across the map), and "clustered" (most intra-map connections are generated within a local neighborhood, with a small admixture of longer-range connections). The "clustered" pattern is most like the one found in cortex. Panels at the left show connection patterns of 40 randomly chosen units, middle panels show a single frame of the unfolding neural dynamics (mpeg movies are available at http://www.indiana.edu/∼cortex/complexity.html), and rightmost panels show spatially averaged activity traces obtained from near the center of the map (circled area). The values for the characteristic path length λ(G), clustering coefficient γ(G), complexity C(X), and total wiring length (l_{wire}) were: $\lambda = 0$, $\gamma = 0$, C(X) = 0.143, $l_{wire} = 0$ ("sparse"); $\lambda = 3.1310$, $\gamma = 0.0076$, C(X) = 0.289, $l_{wire} = 10,807$ ("uniform"); $\lambda = 5.6878$, $\gamma = 0.2637$, C(X) = 0.579, $l_{wire} = 1,509$ ("clustered"), all means of 5 runs. Note that C(X) is highest for the "clustered" network, which shows a rich set of spatiotemporal patterns including waves and spirals. This network also exhibits small-world attributes (low λ(G), high γ(G)) and short wiring length. Modified after Sporns (2004)

and Plenz, 2003; Haldeman and Beggs, 2005; see also Beggs et al., 2007). In neural architectures, critical branching processes give rise to sequences of propagating spikes that form neuronal avalanches. In the critical regime, the branching parameter expressing the ratio of descendant spikes from ancestor spikes is found to be near unity, such that a triggering event causes a long chain of spikes that neither dies out quickly (subcriticality) nor grows explosively (supercriticality). Slice preparations of rat cortex operate at or near criticality, generating neuronal avalanches with a size distribution following a power law (Beggs and Plenz, 2003). Criticality is found to be associated with maximal information transfer and thus high efficacy of neuronal information processing, as well as with a maximal number of metastable dynamical states. These results point to additional important links between structural connectivity patterns and the informational processes carried out within them.

The association of certain kinds of dynamics with particular features of structural connectivity opens up a new computational approach. If we fix key aspects of the dynamics (for example, by enforcing a high value of integration or complexity, or of information integration) and then search for connection patterns that are compatible with this type of dynamics, what relationship, if any, do we find? For example, what kinds of structural connection patterns are associated with high values for integration, complexity or information integration? We used complexity (and other information theoretical measures of functional connectivity, such as entropy or integration) as cost functions in simulations designed to optimize network architectures and found that networks that are optimized for high complexity develop structural motifs that are very similar to those observed in real cortical connection matrices (Sporns et al., 2000a; 2000b; Sporns and Tononi, 2002). Specifically, such networks exhibit an abundance of reciprocal (reentrant) connections, a strong tendency to form clusters and they have short characteristic path lengths. Other measures (entropy or integration) produce networks with strikingly different structural characteristics. While it is computationally expensive to employ most types of nonlinear dynamics in the context of such optimizations, a closer examination of specific connection topologies (sparse, uniform and clustered, or cortex-like) that are simulated as nonlinear systems has shown that the association of small-world attributes and complex functional dynamics can hold for more realistic models of cortical architectures as well (Sporns, 2004; Fig. 7). Thus, high complexity, a measure of global statistical features and of functional connectivity, appears to be strongly and uniquely associated with the emergence of small-world networks (Sporns et al., 2004; Sporns, 2006).

Evolutionary algorithms for growing connectivity patterns have been used in evolving motor controllers (Psujek et al., 2006), networks for path integration (Vickerstaff and DiPaolo, 2005), or in the context of sensorimotor coordination (Seth and Edelman, 2004; Seth, 2005). While many current applications, for example those used in evolutionary robotics, rely on small networks with limited connectivity patterns (due to constraints requiring the convergence of evolutionary algorithms in finite computational time), the gap

to larger, more brain-like networks is rapidly closing. An exciting future avenue for computational research in this area involves the evolution of behaviorally capable architectures that incorporate features of biological organization. Results from this research may ultimately contribute to resolving long-standing controversies such as whether biological evolution inherently tends towards biological structures of greater and greater complexity. Initial studies of evolving connectivity patterns embedded in simulated creatures within a computational ecology (Yaeger and Sporns, 2006) suggest that as econiches become more demanding neural architectures evolve towards greater structural elaboration, elevated levels of plasticity, and with functional activity patterns of higher neural complexity.

7 Conclusion

This review has highlighted recent conceptual and methodological progress in the analysis of complex networks. Some of this progress has been truly impressive, significantly influencing all major areas of life, physical and information sciences. Newly developed tools for complex network analysis are now applied to brain networks, at a pace that appears to be steadily accelerating. Still, despite all this progress, an integrated theory of how brain function emerges from brain connectivity has not yet been achieved. Such a theory will have a dramatic impact. The discovery of principles that link structural, functional and effective connectivity to mind and cognition will lead to a new theoretical foundation for future experimental and theoretical approaches in cognitive neuroscience.

Acknowledgments

O.S. and G.T. were supported by the James S. McDonnell Foundation.

References

Achard, S, Salvador, R, Whitcher, B, Suckling, J, Bullmore, E (2006) A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. J Neurosci 26, 63–72

Albert, R, Jeong, H, Barabási, A-L (2000) Error and attack tolerance of complex networks. Nature 406, 378–382

Albert, R, Barabási, A-L (2002) Statistical mechanics of complex networks. Rev Mod Phys 74, 47–97

Amaral, LAN, Scala, A, Barthelemy, M, Stanley, HE (2000) Classes of small-world networks. Proc Natl Acad Sci USA 97, 11149–11152

Amaral, LAN, Ottino, JM (2004) Complex networks. Eur Phys J B 38, 147–162

Angelucci, A, Lebitt, JB, Walton, EJS, Hipe, J-M, Bullier, J, Lund, JS (2002) Circuits for local and global signal integration in primary visual cortex. J Neurosci 22, 8633–8646

Assisi, CG, Jirsa, VK, Kelso, JAS (2005) Synchrony and clustering in heterogeneous networks with global coupling and parameter dispersion. Phys Rev Lett 94, 018106

Bang-Jensen, J, Gutin, G (2001) *Digraphs: Theory, Algorithms and Applications*. Springer: London etc.

Barabasi, A, Albert, R. (1999) Emergence of scaling in random networks. Science 286, 509–512

Barrat, A, Barthelemy, M, Pastor-Satorras, R, Vespigniani, A (2004a) The architecture of complex weighted networks. Proc Natl Acad USA 101, 3747–3752

Barrat, A, Barthelemy, M, Vespignani, A (2004b) Weighted evolving networks: coupling topology and weight dynamics. Phys Rev Lett 92, 228701

Bassett, DS, Bullmore, E. (2006) Small-world brain networks. Neuroscientist 12, 512–523.

Beggs, JM, Plenz, D (2003) Neuronal avalanches in neocortical circuits. J Neurosci 23, 11167–11177

Bernasconi, C, König, P (1999) On the directionality of cortical interactions studied by structural analysis of electrophysiological recordings. Biol Cybern 81, 199–210

Breakspear, M, Stam, CJ (2005) Dynamics of a neural system with a multiscale architecture. Phil Trans R Soc B 360, 1051–1074

Breakspear, M, Sporns, O, Temby, K, Aquino, K, Van Leeuwan, C (2006) Symbiotic relationship between neuronal dynamics and brain architectures. (in preparation).

Bressler, SL (1995) Large-scale cortical networks and cognition. Brain Res Rev 20, 288–304

Bressler, SL, and Kelso, JAS (2001) Cortical coordination dynamics and cognition. Trends Cogn Sci 5, 26–36

Bressler, SL, McIntosh, AR (2007) The role of context in large-scale neurocognitive network operations. (this volume)

Brodmann, K (1909) Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues. J.A. Barth, Leipzig.

Brovelli, A, Ding, M, Ledberg, A, Chen, Y, Nakamura, R, Bressler, SL (2004) Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality. Proc Natl Acad Sci USA 101, 9849–9854

Büchel, C, Friston, KJ (2000) Assessing interactions among neuronal systems using functional neuroimaging. Neural Netw 13, 871–882

Buzsaki, G, Geisler, C, Henze, DA, Wang, XJ (2004) Interneuron diversity series: Circuit complexity and axon wiring economy of cortical interneurons. Trends Neurosci 27, 186–193

Cajal, SR (1909) Histologie du systéme nerveux de l'homme et des vertebras. Maloine, Paris.

Chartrand, G, Lesniak, L (1996) Graphs and Digraphs. Chapman and Hall, Boca Raton.

Cover TM, Thomas JA: Elements of information theory. New York: Wiley; 1991.

Da F. Costa, L (2004) The hierarchical backbone of complex networks. Phys Rev Lett 93, 098702

Da F. Costa, L, Sporns, O (2005) Hierarchical features of large-scale cortical connectivity. Eur J Phys B 48, 567–573

Darvas, F, Leahy, RM (2007) Functional imaging of brain activity and connectivity with MEG. (this volume)

De Lucia, M, Bottaccio, M, Montuori, M, Pietronero, L (2005) Topological approach to neural complexity. Phys Rev E 71, 016114

Dodel, S, Herrmann, JM, Geisel, T (2002) Functional connectivity by cross-correlation clustering. Neurocomp 44, 1065–1070

Douglas, R, Martin, K (1991) A functional microcircuit for cat visual cortex. J Physiol (Lond.) 440, 735–769

Doyle, JC, Alderson, DL, Li, L, Low, S, Roughan, M, Shalmov, S, Tanaka, R, Willinger, W (2005) The "robust yet fragile" nature of the internet. Proc Natl Acad Sci USA 102, 14497–14502

Eguiluz, VM, Chialvo, DR, Cecchi, GA, Baliki, M, Apkarian, AV (2005) Scale-free brain functional networks. Phys Rev Lett 94, 018102

Eichler, M (2005) A graphical approach for evaluating effective connectivity in neural systems. Phil Trans R Soc B 360, 953–967

Felleman, DJ, and Van Essen, DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1, 1–47

Ferree, TC, Nunez, PL (2007) Primer on electroencephalography for functional connectivity. (this volume)

Friston, KJ (1993) Functional connectivity: the principal-component analysis of large (PET) data sets. J Cereb Blood Flow Metab 13, 5–14

Friston, KJ (1994) Functional and effective connectivity in neuroimaging: A synthesis. Hum Brain Mapping 2, 56–78

Friston, KJ (2002) Beyond phrenology: what can neuroimaging tell us about distributed circuitry? Annu Rev Neurosci 25, 221–250

Friston, KJ, Harrison, L, Penny, W (2003) Dynamic causal modelling. Neuroimage 19, 1273–1302

Friston, KJ (2005) Models of brain function in neuroimaging. Annu Rev Psychol 56, 57–87

Gilbert CD, Wiesel TN (1989) Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. J Neurosci 9, 2432–2442.

Granger, CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438

Gong, P, van Leeuwen, C (2004) Evolution to a small-world network with chaotic units. Europhys Lett 67, 328–333

Kaiser, M, Hilgetag, CC (2004) Edge vulnerability in neural and metabolic networks. Biol Cybern 90, 311–317

Kaiser, M, Hilgetag, CC (2006) Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. PLoS Comput Biol 2, e95.

Kaminski, M, Ding, M, Truccolo, WA, Bressler, SL (2001) Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. Biol Cybern 85, 145–157

Keinan, A, Sandbank, B, Hilgetag, C, Neilijson, I, Ruppin, E (2004) Fair attribution of functional contribution in artificial and biological networks. Neural Comput 16, 1889–1915

Kötter, R, Stephan, KE (2003) Network participation indices: Characterizing component roles for information processing in neural networks. Neural Netw 16, 1261–1275

Haldeman, C, Beggs, JM (2005) Critical branching captures activity in living neural networks and maximizes the number of metastable states. Phys Rev Lett 94, 058101

Harary, F (1969) Graph Theory. Addison-Wesley, Reading, MA.

Hilgetag, CC, Burns, GA, O'Neill, MA, Scannell, JW, Young, MP (2000) Anatomical connectivity defines the organization of clusters of cortical areas in the macaque monkey and the cat. Philos Trans R Soc Lond B Biol Sci 355, 91–110

Hilgetag, CC, Kötter, R, Stephan, KE, Sporns, O (2002) Computational methods for the analysis of brain connectivity, In: Computational Neuroanatomy – Principles and Methods, (Ascoli, G.A. Ed.) Humana Press, pp. 295–335.

Horwitz, B, Tagamets, M-A, McIntosh, AR (1999) Neural modeling, functional brain imaging, and cognition. Trends Cogn Sci 3, 91–98

Horwitz B (2003) The elusive concept of brain connectivity. Neuroimage 19, 466–470.

Humphries, MD, Gurney, K, Prescott, TJ (2006) The brainstem reticular formation is a small-world, not scale-free, network. Proc Roy Soc B Biol Sci. DOI:10.1098/rspb.2005.3354.

Jirsa, VK, Kelso, JAS. (2000) Spatiotemporal pattern formation in continuous systems with heterogeneous connection topologies, Phys Rev E 62, 6, 8462–8465

Jirsa, VK (2004) Connectivity and dynamics of neural information processing. Neuroinformatics 2, 183–204

Kötter, R, Stephan, KE (2003) Network participation indices: Characterizing component roles for information processing in neural networks. Neural Netw 16, 1261–1275

Lee, L, Harrison, LM, Mechelli, A (2003) A report of the functional connectivity workshop, Dusseldorf 2002. NeuroImage 19, 457–465

Liang, H, Ding, M, Nakamura, R, Bressler, SL (2000) Causal influences in primate cerebral cortex during visual pattern discrimination. Neuroreport 11, 2875–2880

Lungarella, M, Ishiguro, K, Kuniyoshi, Y, Otsu, N (2007). Methods for quantifying the causal structure of bivariate time series. *Int. J. of Bifurcation and Chaos* (in press)

Massimi, M, Ferrarelli, F, Huber, R, Esser, SK, Singh, H, Tononi, G (2005) Breakdown of cortical effective connectivity during sleep. Science 309, 2228–2232

Masuda, N, Aihara, K (2004) Global and local synchrony of coupled neurons in small-world networks. Biol Cybern 90, 302–309

McGill, W (1954) Multivariate information transmission. Psychometrika 19, 97–116

McIntosh, AR, Gonzalez-Lima, F (1994) Structural equation modeling and its application to network analysis in functional brain imaging. Hum Brain Mapping 2, 2–22

McIntosh, AR, Rajah, MN, Lobaugh, NJ (1999) Interactions of prefrontal cortex in relation to awareness in sensory learning. Science 284, 1531–1533

McIntosh, AR (1999) Mapping cognition to the brain through neural interactions. Memory 7, 523–548

McIntosh AR (2000) Towards a network theory of cognition. Neural Netw 13, 861–870

McIntosh AR, Rajah MN, Lobaugh NJ (2003) Functional connectivity of the medial temporal lobe relates to learning and awareness. J Neurosci 23, 6520–6528

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. Science 298, 824–827

Milo, R, Itzkovitz, S, Kashtan, N, Levitt, R, Shen-Orr, S, Ayzenshtat, I, Sheffer, M, Alon, U (2004) Superfamilies of evolved and designed networks. Science 303, 1538–1542

Mountcastle, VB (1978) An organizing principle for cerebral function. In: The Mindful Brain, Edelman, G.M. and Mountcastle, V.B. (eds.), pp. 7–50, MIT Press, Cambridge.

Mountcastle, VB (1997) The columnar organization of the neocortex. Brain 120, 701–722

Munk, MH, Linden, DE, Muckli, L, Lanfermann, H, Zanella, FE, Singer, W, Goebel, R (2002) Distributed cortical systems in visual short-term memory revealed by event-related functional magnetic resonance imaging. Cereb Cortex 12, 866–876

Netoff, TI, Clewley, R, Arno, S, Keck, T, White, JA (2004) Epilepsy in small-world networks. J. Neurosci. 24, 8075–8083

Newman, MEJ (2003) The structure and function of complex networks. SIAM Rev 45, 167–256

Nishikawa T, Motter AE, Lai YC, Hoppensteadt FC (2003) Heterogeneity in oscillator networks: are smaller worlds easier to synchronize? Phys Rev Lett 91, 014101

Onnela, J-P, Saramaki, J, Kertesz, Kaski, K (2005) Intensity and coherence of motifs in weighted complex networks. Phys. Rev. E 71, 065103

Paninski, L (2003) Estimation of entropy and mutual information. Neural Comput 15, 1191–1254

Papoulis, A (1991) Probability, Random Variables, and Stochastic Processes. McGraw-Hill, New York.

Passingham, RE, Stephan, KE, Kötter, R (2002) The anatomical basis of functional localization in the cortex. Nature Rev Neurosci 3, 606–616

Pascual-Leone A, Walsh V, Rothwell J (2000) Transcranial magnetic stimulation in cognitive neuroscience – virtual lesion, chronometry, and functional connectivity. Curr Opin Neurobiol 10, 232–237

Paus T (1999) Imaging the brain before, during, and after transcranial magnetic stimulation. Neuropsychologia 37, 219–224

Phillips, CG, Barlow, HB, Zeki, S (1984) Localization of function in the cerebral cortex: past, present and future. Brain 107, 327–361

Pola, G, Thiele, A, Hoffman, K-P, Panzeri, S (2003) An exact method to quantify the information transmitted by different mechanisms of correlational encoding. Network Comput Neural Syst 14, 35–60

Prill, RJ, Iglesias, PA, Levchenko, A (2005) Dynamic properties of network motifs contribute to biological network organization. PLoS Biol. 3, e343

Psujek, S, Ames, J, Beer, RD (2006) Connection and coordination: The interplay between architecture and dynamics in evolved model pattern generators. Neural Comput 18, 729–747

Puchalla, JL, Schneidman, E, Harris, RA, Berry, MJ (2005) Redundancy in the population code of the retina. Neuron 46, 493–504

Roebroeck, A, Formisano, E, Goebel, R (2005) Mapping directed influence over the brain using Granger causality and fMRI. NeuroImage 25, 230–242

Roelfsema, PR, Engel, AK, König, P, Singer, W (1997) Visuomotor integration is associated with zero time-lag synchronization among cortical areas. Nature 385, 157–161

Roxin A, Riecke H, Solla SA (2004) Self-sustained activity in a small-world network of excitable neurons. Phys Rev Lett 92, 198101

Salin, P-A, Bullier, J (1995) Corticocortical connections in the visual system: Structrue and function. Physiol Rev 75, 107–154

Salvador, R, Suckling, J, Schwarzbauer, C, Bullmore, E (2005a) Undirected graphs of frequency-dependent functional connectivity in whole brain networks. Phil Trans R Soc B 360, 937–946

Salvador, R, Suckling, J, Coleman, M, Pickard, JD, Menon, DK, Bullmore, ET (2005b) Neurophysiological architecture of functional magnetic resonance images of human brain. Cereb Cortex 15, 1332–1342

Salvador, R, Achard, S, Bullmore, E (2007) Frequency-dependent functional connectivity analysis of fMRI data in Fourier and wavelet domains. (this volume)

Scannell, JW, Burns, GAPC, Hilgetag, CC, O'Neil, MA, Young, MP (1999) The connectional organization of the cortico-thalamic system of the cat. Cereb. Cortex 9, 277–299

Schneidman, E, Still, S, Berry, MJ, Bialek, W (2003) Network information and connected correlations. Phys. Rev. Lett. 91, 238701

Schneidman, E, Bialek, W, and Berry, MJ (2003) Synergy, redundancy, and independence in population codes. J. Neurosci. 23, 11539–11553

Schreiber, T (2000) Measuring information transfer. Phys Rev Lett 85, 461–464

Seth, A, Edelman, GM (2004). Environment and behavior influence the complexity of evolved neural networks. Adapt Behavior 12, 5–20

Seth, AK (2005). Causal connectivity analysis of evolved neural networks during behavior. Network Comput Neural Systems, 16, 35–54

Shipp, S, Zeki, S (1985) Segregation of pathways leading from area V2 to areas V4 and V5 of macaque monkey visual cortex. Nature 315, 322–325

Song, S, Sjöström, PJ, Reigl, M, Nelson, S, Chklovskii, DB (2005) Highly nonrandom features of synaptic connectivity in local cortical circuits. PLoS Biology 3, e68

Sporns, O, Tononi, G, Edelman, GM (2000a) Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices. Cereb. Cortex 10, 127–141

Sporns, O, Tononi, G, Edelman, G (2000b) Connectivity and complexity: the relationship between neuroanatomy and brain dynamics. Neural Netw 13, 909–922

Sporns, O., Tononi, G (2002) Classes of network connectivity and dynamics. Complexity 7, 28–38

Sporns, O (2002) Graph theory methods for the analysis of neural connectivity patterns. Kötter, R. (ed.) Neuroscience Databases. A Practical Guide, pp. 171–186, Klüwer, Boston, MA.

Sporns, O, Zwi, J (2004) The small world of the cerebral cortex. Neuroinformatics 2, 145–162

Sporns, O (2004) Complex neural dynamics. In: Coordination Dynamics: Issues and Trends, Jirsa, V.K. and Kelso, J.A.S., (eds.), pp. 197–215, Springer-Verlag, Berlin.

Sporns, O, Kötter, R (2004) Motifs in brain networks. PLoS Biology 2, 1910–1918

Sporns, O, Chialvo, D, Kaiser, M, Hilgetag, CC (2004) Organization, development and function of complex brain networks. Trends Cogn Sci 8, 418–425

Sporns, O (2005) Principles and methods in the analysis of brain networks. In: Reeke, G.N., Poznanski, R.R., Lindsay, K.A., Rosenberg, J.R. and Sporns, O. (eds.) Modeling in the Neurosciences. From Biological Systems to Neuromimetic Robotics, 2nd edition, pp. 599–612, CRC Press, Boca Raton.

Sporns, O, Tononi, G, Kötter, R (2005) The human connectome: A structural description of the human brain. PLoS Comput. Biol. 1, 245–251

Sporns, O (2006) Small-world connectivity, motif composition, and complexity of fractal neuronal connections. BioSystems 85, 55–64

Srinivasan, R, Russell, DP, Edelman, GM, Tononi, G (1999) Increased synchronization of neuromagnetic responses during conscious perception. J Neurosci 19, 5435–5448

Stam, CJ (2004) Functional connectivity patterns of human magnetoencephalographic recordings: A 'small-world' network? Neurosci Lett 355, 25–28

Stephan, KE, Friston, KL (2007) Models of effective connectivity in neural systems. (this volume)

Striedter, GF (2005) Principles of Brain Evolution. Sinauer, Sunderland, MA.

Strogatz, SH (2001) Exploring complex networks. Nature 410, 268–277

Swanson, LW (2003) Brain Architecture. Oxford University Press, Oxford.

Tanigawa, H, Wang, Q, Fuujita, I (2005) Organization of horizontal axons in the inferior temporal cortex and primary visual cortex of the macaque. Cereb Cortex 15, 1887–1899

Tononi, G, Sporns, O, Edelman, GM (1994) A measure for brain complexity: relating functional segregation and integration in the nervous system. Proc Natl Acad Sci USA 91, 5033–5037

Tononi, G, Edelman, GM (1998) Consciousness and complexity. Science 282, 1846–1851

Tononi, G, Edelman, GM, Sporns, O (1998a) Complexity and coherency: Integrating information in the brain. Trends Cogn Sci 2, 474–484

Tononi, G, McIntosh, AR, Russell, DP, Edelman, GM (1998b) Functional clustering: Identifying strongly interactive brain regions in neuroimaging data. Neuroimage 7, 133–149

Tononi, G, Sporns, O, Edelman, GM (1999) Measures of degeneracy and redundancy in biological networks. Proc Natl Acad Sci USA 96, 3257–3262

Tononi, G (2001) Information measures for conscious experience. Arch Ital Biol 139, 367–371

Tononi, G, Sporns, O (2003) Measuring information integration.BMC Neuroscience 4, 31

Tononi G (2004) An information integration theory of consciousness. BMC Neuroscience, 5, 42

Ungerleider, LG, Mishkin, M (1982). Two cortical visual systems. in Analysis of visual behaviour, Ingle, D. G., Goodale, M. A. and Mansfield, R. J. Q., ed., MIT Press, Cambridge, MA, pp. 549–586

Van Essen, DC, Maunsell, JHR (1983) Hierarchical organization and functional streams in the visual cortex. Trends Neurosci 6, 370–375

Varela F, Lachaux JP, Rodriguez E, Martinerie J (2001) The brainweb: phase synchronization and large-scale integration. Nat Rev Neurosci 2, 229–239

Vickerstaff, R, Di Paolo, EA (2005) Evolving neural models of path integration. J Exp Biol 208, 3349–3366

Von Stein, A, Rappelsberger, P, Sarntheim, J, Petsche, H (1999) Synchronization between temporal and parietal cortex during multimodal object processing in man. Cereb Cortex 9, 137–150

Watts, DJ, Strogatz, SH (1998) Collective dynamics of 'small-world' networks. Nature 393, 440–442

Watts, DJ (1999) Small Worlds. Princeton University Press, Princeton, NJ.

Watts, DJ (2003) Six Degrees. The Science of a Connected Age. W.W. Norton, New York.

Yaeger, L, Sporns, O (2006) Evolution of neural structure and complexity in a computational ecology. In Rocha, L. et al. (eds.) *Artificial Life X*. Cambridge, MA: MIT Press (in press)

Young, MP (1993) The organization of neural systems in the primate cerebral cortex. Proc R Soc Lond B 252, 13–18

Zeki, S, Shipp, S (1988) The functional logic of cortical connections. Nature 335, 311–317

Zeki, S (1993) A Vision of the Brain. Blackwell, London

Zhigulin, VP (2004) Dynamical motifs: Building blocks of complex dynamics in sparsely connected random networks. Phys Rev Lett 92, 238701

Anatomical Concepts of Brain Connectivity

Rolf Kötter

Section Neurophysiology & Neuroinformatics, Dept. of Cognitive Neuroscience, Radboud University Medical Centre, Nijmegen, The Netherlands.
Inst. of Anatomy II and Vogt Brain Research Institute, Heinrich Heine University, Düsseldorf, Germany.

Introduction

The anatomy of brain wiring is often taken as an established and permanent body of knowledge, sometimes considered as a 'ground truth', which supposedly provides a secure basis for validating structural and functional imaging data, for building neural network models, or for constraining inferences on causal mechanisms generating brain activity.

Macroscopical brain anatomy changes only slowly with ontogenetic development. Also, at any stage morphological data are less variable than most functional measurements. Nevertheless, there is considerable dynamics at the microscopical scale. Also, observer- and method-dependence determine which structural features are unanimously recognized and which others depend critically on sensitivity and specificity, or are inferred on the basis of pre-existing knowledge. Many structural features vary between species or within species or with plasticity and learning. There are also structural features that lack a known functional correlate, or where the structure and the function are not as tightly coupled as we would expect them to be. This variety of observations illustrates that it is important to know the circumstances and limitations of anatomical observations.

Here I present anatomical methods used for establishing structural connectivity in the brain and discuss some prerequisites that have to be fulfilled to ensure that the anatomical data on brain connectivity support the inferences that we like to base upon them.

Types of Anatomical Connectivity and Methods for their Detection

The most fundamental principle to be aware of when referring to 'anatomical' or 'structural' connectivity is that these terms may relate to several different morphological entities depending on the level of detail and the method of measurement (Fig. 1).

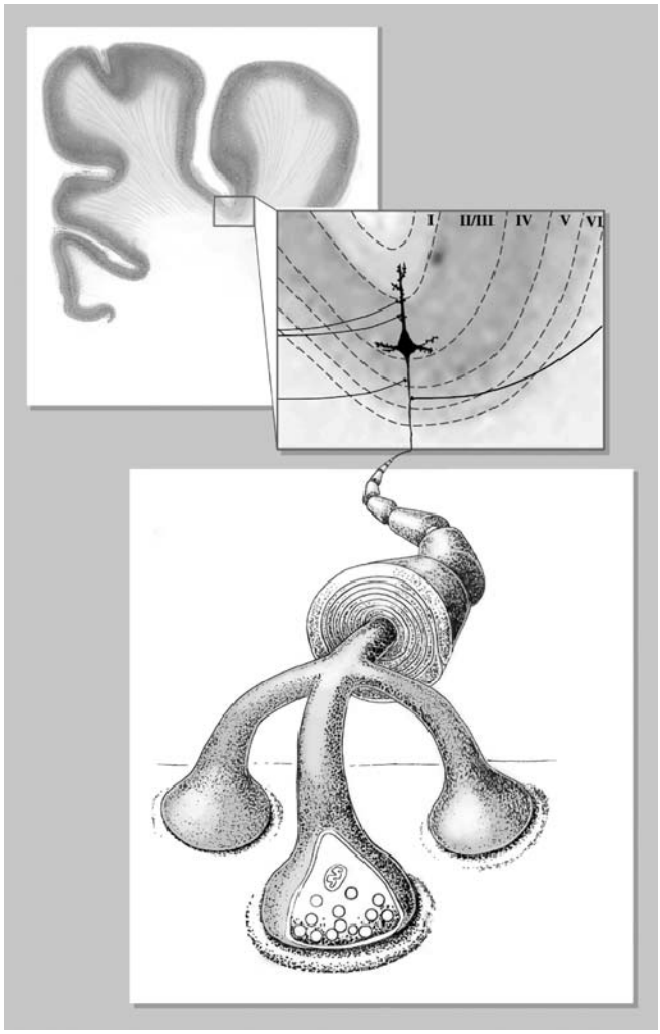


Fig. 1. Levels of anatomical detail in the description of brain connectivity. Top: white matter tracts and interregional projections; middle: neuron types and microcircuits; bottom: axon terminals with synapses

Neuronal Connectivity

Functionally meaningful connections in the brain are generally made by neurons (that is ignoring the potential role of glia and other cells in direct information processing as opposed to their support of neuronal functions). Therefore, the basic unit of a connection in the brain is the contact from one neuron to another, which is characterized by the presence of a specific structure: the neuronal synapse. Synapses have a defined morphological correlate, which can be observed by transmission electron microscopy: For chemical synapses

it requires the demonstration of presynaptic vesicles, a postsynaptic density, and a synaptic cleft of about 20 nm width (see Fig. 2; Gray 1959; Jones 1975). Thus, the demonstration of anatomical connectivity at the neuronal level strictly speaking requires electron microscopic (EM) studies.

At this level of detail it is currently only possible to examine small specimens in the µm to mm range, which in many nervous systems does not even comprise the entire extent of two communicating neurons. The enormous difficulty to reconstruct and label structures in serial electron micrographs explains why the nematode worm Caenorhabditis elegans with a length of about 1 mm still remains the only organism whose nervous system has been fully visualized at the EM level. Although the connectivity of its 302 neurons is largely determined (White et al. 1986), even in C. elegans the morphological details of some neuronal connections have not yet been worked up completely (www.wormbase.org). Electron microscopic reconstructions of an entire cortical column with its roughly 10,000 neurons are being considered, but have not yet been seriously attempted mainly due to the limitations of automatic reconstruction and analysis of corresponding EM structures in adjacent sections (Fiala 2002; Denk and Horstmann 2004).

The best approximation at the light-microscopical level to the demonstration of a synapse is the visualization of terminal boutons ('button'-like axonal swellings of 0.5–3 µm diameter). Boutons can be seen most clearly in single neurons after intracellular filling, but require confocal microscopy or very thin sections if their number and their presumed synaptic partners are to be established. Still today the most serious researchers prepare their light-microscopical sections such that selected regions of the specimens can be

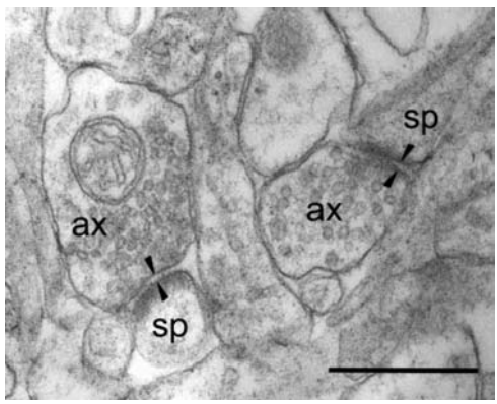


Fig. 2. Transmission electron microscopic image of synapses with pre- and postsynaptic components as the morphological demonstration of the apparatus for electrochemical signal transmission between nerve cells. The picture shows axon terminals (ax) making Gray type I (usually cortical excitatory) synapses (arrows) on spines (sp) of medium-sized spiny neurons in the striatum. Scale bar: 0.25 µm

subsequently cut out, re-sectioned and investigated in more detail under the electron microscope.

While synapses and boutons are hallmarks of the connections between neurons, these do not unambiguously establish the type and somatic location of the pre- and postsynaptic neurons. The morphological neuron type is determined as a first approximation by its soma location, the somatodendritic shape, the presence of dendritic spines, the pattern of axonal termination, and the combinations of these features. Not all necessary determinants may be observed in the same preparation, which leads to incomplete morphological classifications such as 'non-pyramidal cell' or 'aspiny interneuron'. Although we now have many selective cytochemical markers allowing us to classify individual cells, this classification tells us little about the precise connectivity of this cell beyond the statistics associated with its general type. Using recent molecular approaches to classify neurons based on the expression of individual genes (see www.gensat.org) it may be possible in the future to deduce the type of connectivity of individual neurons from the presence of more specific genetic markers.

The most detailed studies of structural and functional neuronal connectivity so far have been performed using intracellular recordings from nearby neurons within a cortical column or a subnucleus. Usually the cells are visually targeted under the microscope and then individually impaled with glass pipettes containing a hydrophilic diffusible marker of low molecular weight, such as biocytin. This marker spreads within minutes throughout the cytoplasm of the impaled cell and labels the spatially extended neuropil, which can be observed by fluorescence microscopy (MacDonald 1992). Neuron types in the cerebral cortex are therefore widely recognized although it remains unclear exactly how many types there are and how they would be classified best.

Simultaneous targeting of several neurons has demonstrated that the probability of direct anatomical and functional connections between cells decreases rapidly with the distance between their cell bodies. Neurons that are monosynaptically connected show similar connectivity and electrophysiological properties (Yoshimura et al. 2004). In general, however, the statistical likelihood of finding two directly connected neurons at longer distances in the cerebral cortex is zero most of the time. This statement may not seem intuitive given the calculation that on the average any two cortical neurons are connected via three consecutive synapses between them (Braitenberg and Schüz 1998). It highlights, however, the crucial role of the type of connectivity patterns in networks where each neuron makes large numbers of connections in absolute terms, but contributes only a tiny fraction relative to the total number of such connections. The role of some such patterns in cortical networks is considered in the chapter by Sporns and Tononi in this volume.

Regional Connectivity

Moving from connections between individual neurons to connections between brain regions, the most widespread and valuable method delivering detailed information about directed long distance connections is neuroanatomical tract tracing (for reviews see Sawchenko and Swanson 1981; Köbbert et al., 2000; Wouterlood et al. 2002). The general approach comprises now a vast range of substances with the common feature that they are taken up by neurons and spread along their projections, where the label then can be visualized. Some substances are directly inserted intracellularly and therefore suitable for tracing of individual neurons. Most of the tracer substances, however, are applied extracellularly to the living tissue by pressure injection, iontophoresis or mechanical insertion. Most of them are actively incorporated through the neuronal membrane and transported in the cytoplasm to reveal the distant location of cell bodies (fast retrograde transport) or axonal terminals (fast and slow anterograde transport). A wide variety of substances are used for tract tracing, which differ in the direction and speed of transport, completeness and intensity of staining, sensitivity and persistence, and the mode of their detection. Among the best tracer substances are plant lectins, such as Phaseolus vulgaris leucoagglutinin (PHA-L), which bind to glycoconjugates on the neuronal membrane and are rapidly internalized and transported both antero- and retrogradely (Gerfen and Sawchenko 1984). Lectins are being used either alone or coupled to retrograde fluorescent tracers, horseradish peroxidase (HRP) or cholera toxin subunit B. Radioactive tracers, mainly tritiated amino acids (TAA), are hydrophilic tracers that are readily incorporated and transported anterogradely even across the synaptic cleft. They have now been largely replaced by dextran amines, which are easier to handle and to detect. For transneuronal tracing, viruses, such as the rabies and herpes virus, are being tried with the limitation that transgression to subsequent neurons may involve extrasynaptic sites (Boldogkoi et al. 2004; Hoshi et al. 2005). Problematic are degeneration studies after physical or toxic destruction of neuronal somata or axon dissection, with subsequent visualization of anterograde (Wallerian), retrograde, or transneuronal degeneration signs (Nauta 1957). The signs of degeneration may be difficult to detect, and large lesions tend to affect fibres of passage. Lipophilic carbocyanine dyes, such as DiI and DiO, spread by lateral diffusion within the cell membrane. Besides in living tissue these can also be employed in fixed tissue, notably for post-mortem in the human brain (Honig and Hume 1989). Although the speed of diffusion increases with temperature, the range of connections identified in fixed tissue is limited to distances of less than 1 cm and the quality of the images is comparatively poor. Thus post-mortem tracing is currently not suitable for tracing of long-distance connections in the human brain.

In *in vivo* tracing studies, the tracer is actively transported and maximally concentrated in the axon terminals (anterograde transport) or the cell body (retrograde transport) after appropriate survival periods of days to weeks.

The number of retrogradely marked cell bodies can be counted and the proportions in different afferent structures be compared. Axon terminals are hard to quantify leading to density measures in the best case and, more commonly, to rough rankings from sparse to dense labelling. In layered structures, such as the cerebral cortex, the laminar distribution of label at the site of transport can be observed, whereas the laminar position at the application site is often doubtful due to the size of the application and its diffusion halo, where additional uptake may have occurred. Furthermore, tracing studies provide information primarily about the sites of application and labelling after transport, whereas the course of the axonal projections is usually not well observed and hardly ever spatially reconstructed. This limitation results in the curious situation that we know much about which regions are connected by direct and directed axonal projections, but not by what route.

Several other anatomical methods have been tried to gain information about structural connectivity in the brain. Staining for myelin, particularly in fresh tissue using the Gallyas or Heidenhain-Woelcke stains, reveals the presence of myelinated axons predominantly in the white, but to some degree also in the grey matter. Myelin is ubiquitous in the white matter, which makes it impractical to identify individual fibres or to specify the connected sources and targets. Modified de-staining techniques can leave small sets of fibres visible so that their course can be followed. The difficulty of staining specific tracts limits the use of myelin stains mainly to the parcellation of brain structures on the basis of differential myelin densities.

Other methods, such as polarization microscopy, specify the mean orientation of measure fibres in small regions and allow the spatial reconstruction of local fibre bundles. Obtaining information about the three-dimensional arrangement of fibre bundles is useful to classify major tracts in the white matter, but it remains the uncertainty whether all fibres follow the overall course of the bundle. In some cases, such as the medial forebrain bundle, the bundle can be more appropriately regarded as a conduit, which groups fibres together for some part of their course with individual fibres coming in and leaving at different levels, and possibly not a single one actually continuing throughout the whole length of the bundle. By contrast, it is reasonable to assume that some fibre tracts contain a homogeneous set of fibres that originate from and terminate in a single structure (e.g. the geniculo-cortical tract in the primate visual system).

It may not be obvious to address diffusion weighted brain imaging in a section on anatomical technique. However, if one accepts imaging of calcium- or voltage-sensitive dyes as an electrophysiological method then it seems equally valid to consider the contribution of magnetic resonance imaging (MRI) to the identification of brain structure and anatomical connectivity. This inclusion is particularly relevant since diffusion weighted brain imaging is rapidly gaining the status of a versatile substitute for missing anatomical studies and is frequently being interpreted without proper validation against direct anatomical data.

Diffusion weighted magnetic resonance imaging (dMRI; more specifically diffusion tensor imaging, DTI) relies on special pulse sequences, which elicit signals that reflect the orientation of diffusion processes in soft tissue, specifically of water molecules in the brain (Le Bihan 2003). The Brownian motion of water molecules is strongly constrained by myelinated fibre tracts, which hinder any motion across them. The signals therefore indicate two main features: the deviation from randomness of diffusion within each measured voxel, which is expressed by the fractional anisotropy (FA), and the three-dimensional orientation of the diffusion tensor or the corresponding probability functions. Comparing the tensors in adjacent voxels one can concatenate the most consistent orientations ('tractography') and visualize them as lines, which correspond to major white matter fibre tracts in the human and macaque brain (Tuch et al. 2005). This abbreviated description may suffice to understand that dMRI does not directly show fibre tracts (see chapter by Alexander and Lobaugh in this volume), and that the visualization depends to a great extent on the parameters of the reconstruction algorithm. In addition, the voxel sizes are comparatively large at about 6 mm for in vivo imaging with measurement durations not exceeding 1 hour. Tractography cannot disambiguate the many possible geometric constellations within a voxel that lead to the same signal as, for example, crossing and 'kissing' of fibre tracts. Thus, dMRI results are based on major assumptions, which need to be thoroughly validated before they could be equated with directly demonstrated anatomy. Diffusion-weighted MRI is the method of choice for longitudinal in vivo studies of fibre tracts in the whole brain, whereas invasive tract tracing remains the gold standard for identifying selected neuronal connections at a single point in time. The two methods are complementary in so many respects that it is not at all obvious how they could be employed for direct cross-validation.

The characteristics of tract tracing experiments and diffusion weighted imaging are compared in Table 1.

At this time it appears that cross-validation requires a number of additional approaches:

- Combination of traditional tract tracing with corresponding paramagnetic tracers that can be visualized by in vivo whole brain imaging to demonstrate that the two techniques show corresponding results in the same animal (e.g. Saleem et al., 2002).
- In vivo imaging of paramagnetic tracers in comparison with dMRI in the same species to evaluate the reliable demonstration of non-trivial fibre tracts.
- Comparison of dMRI with other indirect measures of connectivity (such as the functional connectivity measured as ongoing "resting state" activity) to evaluate which one is the best predictor of anatomical connectivity.

Before these relationships are established and the applicability of the necessary animal studies to other species, particularly the human brain, is shown, the interpretation of indirect in vivo imaging results as demonstrating anatomical connectivity is not justified.

Table 1. Comparison of characteristic features of tract tracing and diffusion-weighted imaging

Tract tracing	Diffusion-weighted imaging
microscopic resolution	low spatial resolution (several mm)
any species (limited by ethics)	applicable to large brains (humans)
fibres quantifiable	surrogate myelin measures
long history and known validity	unclear validity
invasive animal experiment	non-invasive, applicable to humans
post-mortem assessment	longitudinal studies possible
few injections per brain	entire brain imaged
few population studies done	population studies usually done
3D reconstructions rarely done	registered in 3D coordinates
fibre origin + destination	shows fibre course, no direction
applicable to any brain structure	limited to white matter
can be combined with histology	can be combined with struct./fMRI

Mapping Neuronal and Regional Connectivity

For understanding the connectional architecture of the brain it would be very helpful if neuronal and regional connectivity could be related to each other. Unfortunately, their relationship is not straightforward.

When studying neurons in brain sections the dendritic tree and even most of the local axon arborizations can be assessed (Staiger et al. 1999). It is not known, by contrast how many of the distant interregional projections have been severed. It is exceedingly rare that individual axons have been followed and reconstructed beyond the direct vicinity into other brain regions (e.g. Rockland et al. 1999; Zhong et al. 2006). From the available studies, however, important general rules have been gleaned, which are of great significance for our understanding of cortical organization. Such general rules state that interregional projections originate from neurons in supra- and infragranular layers, that the projection neurons (emanating intrahemispheric association fibres, contralateral commissural fibres and extracortical projection fibres) are pyramidal cells, and that the size of the pyramid-shaped somata is roughly related to the length of their axons. Since pyramidal cells invariably use excitatory amino acids (mainly glutamate) as their transmitter, it is reasonable to infer that all interregional projections are excitatory. The effect of one cortical area on another, however, depends not only on the type of projection neuron but also on the type of neuronal targets and their dynamics: Functionally, association fibres appear to have predominantly excitatory effects on most cortical regions, whereas commissural fibres tend to evoke contralateral inhibition (e.g., Ferbert et al. 1992).

There remains the question of what it means when we say that one brain region projects to another? Here, brain regions, in particular cortical areas, are conceptualized as homogeneous units whose characteristics are invariant along major spatial dimensions. Nevertheless, the criterion of homogeneity

depends on the particular feature studied with the result that there are many different definitions of brain regions, which vary in their level of detail and precision. This variability is comparable to the variability in neuron classifications at the cellular level, and the uncertainties resulting from the two sets of variability are independent and thus not constrained by mutual comparisons. As a result of this enormous variability it is difficult to obtain precise statements on the neuronal underpinnings of interregional connectivity. For example, along the vertical dimension of the cerebral cortex different layers of the same cortical region have long been known to reach different intra- and subcortical targets; the horizontal dimension along *neurons in the centre and in*, the periphery of cortical columns have different projection preferences among neighbouring columns with different functional consequences. Such intraregional differences need to be taken into account at the level of interregional projections.

Current interregional tract tracing experiments provide some information about vertical intraregional differentiation referring to the laminar origin and termination of interregional fibres in the cerebral cortex: In analogy to the ascending thalamocortical fibres, which terminate predominantly in layer IV of their cortical target regions, corticocortical fibres that target preferentially layer IV of another region are also referred to as ascending or feedforward connections. By contrast, the axon terminals of reciprocal pathways usually avoid layer IV and are consequently named descending or feedback connections (Rockland and Pandya 1979; Felleman and Van Essen 1991). Although the clarity of laminar termination preferences varies with the presumed hierarchical distance between two connected areas, the most obvious ascending termination patterns arise from supragranular neurons, whereas infragranular neurons contribute predominantly to descending projection patterns (Barone et al. 2000). A third type of columnar termination pattern across all layers is attributed to so-called 'lateral' projections between cortical regions that are supposed to belong to similar levels in the processing hierarchy.

In contrast to the laminar preferences, the neuronal preferences of axon terminals are largely unknown. For example, it is not known whether interregional projections contact excitatory or inhibitory interneurons or projection neurons in their target regions. Knowing the identity of those target cells would be very important for understanding the functional impact of interregional projections (see difference in functional impact between association fibres and commissural fibres mentioned above). One may even go as far as measuring the volume of axonal boutons of region-specific projections to obtain an estimate of their functional impact on distinct cortical layers (Germuska et al. 2006).

It is also relevant to obtain an estimate of the specific processing of extrinsic versus intrinsic signals, and of the likelihood and timing of further transmission to subsequent brain regions. It has been estimated that even in primary sensory regions thalamocortical afferents form only a small proportion of the total number of synapses, even in the dominant target layer IV in

the cerebral cortex. Whether networks of intrinsic granule cells act to boost thalamocortical signals is a matter of speculation although they certainly possess the required connectional features (e.g. Schubert et al. 2003). It is likely that thalamocortical inputs contribute only a fraction of the input required to activate a cortical brain region. Thus an input-driven mean-field model of cortical activity spread (e.g. Kötter et al. 2002) is a gross simplification applying to the situation where the cortical system is ready to respond to such inputs. To improve our understanding of whether certain inputs lead to a significant change in cortical activity we depend crucially on more detailed data on their cellular targets. This may serve as an example that the characteristics of local microcircuits will influence the models of global activity patterns in the cerebral cortex.

Assembling Connectivity Matrices

Interregional connectivity patterns of the brain are often represented as connectivity matrices of size n^2, where n is the number of brain structures. While the matrix representation is a very concise notation and widely used it can be very differently interpreted. It is commonly assumed that every row or column entry is a single entity, which represents a unique part of the brain. The union of all entities would then be equivalent to a coherent part of the brain such as the visual cortex or the set of thalamo-cortical projections. While such sets of brain regions more often than not represent only a fraction of the brain system of interest, the designation of individual locations may not be unique, for example, when subdivisions are simultaneously listed with a supra-ordinate brain structure or when overlapping entities occur in the same matrix. Several such mapping issues, which are relevant to understanding the scope and reliability of structural and functional connectivity data have been discussed in detail by Kötter and Wanke (2005)

Analyses of connectivity matrices may be influenced by the resolution applied: At the level of lobes the cerebral cortex is a completely connected network, using area granularity the connection density declines to about 50%. Finer divisions produce sparser matrices because of the absence of detailed data but probably also in reality since long-range axons show locally clustered branching patterns leading to increasing sparseness until the resolution matches the branch cluster size. Therefore, when analyzing connectivity matrices it is relevant to note the absolute and relative size of the entities and to consider alternative partitioning schemes as controls.

Over the last 15 years tract tracing data have been collated in several non-human species: rats, cats, monkeys (mainly macaques). Some of these have become legacy data. In macaque monkeys, the first comprehensive review of connectivity within the visual system and within the sensorimotor system was published by Felleman and Van Essen (1991). Connectivity matrices resulting from this study have been published and analyzed by others numerous times (see e.g. Sporns and Tononi in this volume). Young added additional data from

his reading of the literature and analyzed for the first time a matrix comprising almost the entire cerebral (neo-)cortex at the regional level (Young 1993). Improvements in databasing technology and coordinate-independent brain mapping have subsequently led to a systematic effort in collating tracing data from the entire forebrain in macaques with about 40,000 entries in many different parcellations schemes (www.cocomac.org; Stephan et al. 2001; Kötter et al. 2004). Several specialized regional matrices have been published and analyzed subsequently (e.g. Bozkurt et al. 2001, 2002; Kötter and Stephan 2003).

Related efforts were made to gain an overview of the regional cortical connectivity (Scannell et al. 1995) and thalamocortical connectivity (Scannell et al. 1999) in the cat. The cortical regions included a large extent of allocortex. Burns collected and analyzed regional connectivity of the allocortex and the hypothalamus of the rat (e.g. Burns and Young 2000).

All these efforts relied on data from published anatomical tracing studies. Although these have contributed much to our understanding of cortical organization they are lacking detail addressing important issues: Quantitative and laminar data are rare, and-where available-they do not cover much of the cortex. For example, many data on connections between visual areas in the macaque have been specified in terms of their laminar origin and termination, but they the anatomical density of fibres ("strength") has not received the same amount of attention. The situation is almost the opposite in the sensorimotor system where a group of researchers around Rizzolatti have performed extensive quantifications of connection density. But even fundamental data on the gender of the animals or on the identity of the investigated hemisphere are frequently missing even though they must have been known at the time of the experiments. Whether his shortcoming results from the opinion that such differences are not important or so small that they cannot be demonstrated in a single study or in a single laboratory, it now hampers the insights that could be gleaned from meta-analyses in large data collations. While there is still much information to be gained from investigating tracing data, this does require more detailed attention to the available data and suitable methods for analyzing them.

While macaques are a well investigated genus with particular relevance to the human brain, the wide availability of rodents has resulted in more detailed investigations at the columnar and cellular levels. This bears the promise to bridge levels and to understand the relationship between them. Unfortunately, not much corresponding efforts have been made to match investigations at the different levels.

Summaries of connectivity data suitable for generating matrices at the cellular or laminar levels are rare (see e.g. Häusler and Maass 2007; Bannister 2005; Thomson and Bannister 2003). Corresponding morphological studies often describe individual cases whereas connectivity matrices show cell types, which depend on the classificatory scheme applied. For example, whether every supragranular pyramidal cell has the same cellular or laminar targets or whether the correct number of fundamental inhibitory neuron types has been distinguished, such issues remain a matter of controversy.

There is some hope that characterizations at the molecular level could bring to light a fundamental underlying principle that would motivate a meaningful objective classification. So far, multivariate classifications of neurons based on mRNA or peptide expression (e.g. Markram et al. 2004; Sugino et al. 2006) or genetic constructs for visualizing cells expressing a specific gene product (Hatten and Heintz 2005) have provided exciting new ways of classifying neurons, but they have not resulted in a unified objective scheme.

Thus, the classification of individual cells, similar to the classification of brain structures, relies to a large degree on subjective experience and group consensus. Objective quantitative and universally recognized measures are still elusive.

Last not least, there are simpler animals where the whole nervous system has been targeted. Probably the simplest vertebrate whose motor behaviour has been extensively analyzed to the level of repetitive circuits is the lamprey (e.g., Grillner and Wallen 2002). Physiologically motivated studies that include the anatomy of cellular circuitry are being carried out in invertebrates, such as the leech (e.g., Zheng et al. 2007). Finally, there have been the electron microscopic studies of the whole organism in C. elegans mentioned in the section on neuronal connectivity above (White et al. 1986).

Perspectives in Demonstrating Anatomical Connectivity

Labelling all Synaptic Connections of a Single Neuron

Labeling of "one neuron with a substance that would then clearly stain all the neurons connected to it, and no others" (Francis Crick) has been a sort of holy grail in the field of neural circuitry, and many attempts to achieve this have failed. Wickersham et al. (2007) have developed a strategy which uses a modified rabies virus whose genome is missing a gene required to make infectious viral particles. It would thus replicate within a neuron but spread no further because it is missing an envelope glycoprotein that it requires to leave the cell. They further modified the virus so that it would only be able to infect cells expressing a receptor for an avian virus called TVA. By expressing TVA, as well as the missing envelope gene, it was possible to initiate viral infection in that one cell and to produce infectious viral particles. By virtue of complementation of the missing gene, these viral particles then spread to and labelled all of the cells that were retrogradely connected to the first neuron. But then the virus could spread no further because the infected neurons did not express the missing envelope gene. Since the genetic modification of the rabies virus included also a sequence expressing Green Fluorescent Protein (GFP), the spread of the virus could be visualized under the fluorescence microscope. This method was applied to pyramidal cells in cortical slice cultures and is compatible with in vivo physiological assays that will allow the visual responses of single neurons to be directly correlated with their detailed connectivity. If successful this method will be invaluable for teasing apart the detailed circuitry of the brain.

Molecular Approaches to Study Neuronal Cell Types and their Connections in Mice

Genomic approaches can be used to visualize specific neuronal cell types including their axonal projections. Most of the mouse genome is cloned as large fragments (150–200 kilobases in size) in so-called bacterial artificial chromosomes (BACs) — vectors that can be easily propagated inside cells. The advantage of BACs is that they are usually large enough to carry both the coding region of a gene and its regulatory sequences, which determine where and when the gene is expressed. Using a library of BACs, one can systematically replace the coding regions of different genes with a 'reporter' sequence that encodes a fluorescent protein (enhanced green fluorescent protein, or EGFP). By injecting a modified BAC into a mouse egg, one can generate transgenic animals in which we can study the spatiotemporal expression patterns of each gene in the central nervous system. This approach is currently being pursued in the Gene Expression Nervous System Atlas (GENSAT) BAC Transgenic Project (www.gensat.org), which has already provided data on some 150 genes, but aims to provide detailed expression maps for many more (Gong et al. 2003). The BAC method still faces several problems such as specificity when used with very large genes whose regulatory sequences do not fit into a single BAC, or sensitivity where several copies of a BAC are required to make the fluorescent signal strong enough to detect it.

Where EGFP is expressed throughout the entire neurons including the axon, this approach has the potential to trace projections of neurons that are characterized by their molecular genetic properties rather than by their morphological and physiological appearance. For example, the GENSAT project has revealed cell classes specific to individual brain regions and to the different layers of the cerebral cortex. Visualizing their axonal projections in 3D space will be required to map the connectivity patterns of genetically characterized neurons.

Large Scale Systematic Tract Tracing in the Rodent Brain

Tract tracing studies require a lot of expertise and are tedious to perform. Therefore they are usually hypothesis-driven experiments using the method that appears most suitable to address the specific question. This makes it difficult to compare tracing studies even within the same species. This limitation during could be overcome, however, if the studies could be automated both the experiments and the subsequent analyses. For a large-scale systematic tract tracing study of a substantial part of the brain a number of requirements must be met: 1) Since the transport of at most a few tracer deposits can be analyzed in the same brain there must be a large enough population of brains to study their circuitry including its possible variability. 2) The tracer injections should be made in unbiased locations so that an objective estimate of the circuitry is obtained. Ideally, the entire brain would be divided into voxels of

an adequate size and an injection be made in every single one of them. 3) The further processing should be automated as much as possible such that the results can be compared across individual brains. This requires standardized protocols for tissue processing, label detection and 3D reconstruction.

Depending on the size of the injections and the size of the brain one would have to deal with about 100,000 locations even with white matter excluded. Given that each location would need to be studied about three times and that one would want to exclude any ambiguity from injecting several locations at once, this project would seem unlikely to be done in the near future even by large consortia, or by the Allen Brain Institute, which mapped the expression patterns of almost the entire genome in the mouse brain.

Assembling the Connectivity Matrix of the Human Brain

Since genetic approaches are not an option to chart the connectivity of the human brain how could we make progress here? As the human brain is comparatively large we have the advantage of being able to use non-invasive imaging techniques to obtain structural and functional images with a resolution in the millimeter range. As defined above this would allow us to detect regional connectivity but not neuronal connectivity, and at present we have to rely on indirect measurements indicating anatomical connections. Creating the connectivity matrix of the human brain, however, is a worthwhile endeavour with great importance for cognitive neuroscience and neuropsychology. It would form a unique neuroinformatics resource for reference in a variety of fields, which would thereby make closer contact and maybe allow cross-linking of studies that were conceived with a very specific question in mind but have wider implications.

As we suggested previously (Sporns et al. 2005) assembling the connectivity matrix of the human brain would require a multi-modal approach, starting from the macroscopic level using structural and functional in vivo imaging techniques. It would eventually make contact with the mesoscopic level of cortical columns and layers, information that is conceived to be generic to various locations and would thus be mapped from a certain zoom level onwards independent of the exact coordinates. Cellular information (including animal to the mesoscopic level data) could then be linked to maintaining strict transparency of the origin of the data. In pursuing this goal a series of steps would have to be performed, which build on the issues faced with validating DTI as predicting anatomical connections (see section on Regional Connectivity above).

- Step 1: Probabilistic tractography of diffusion-weighted imaging data starting with thalamocortical tracts followed by U-fibres of all cortical regions ultimately resulting in a voxel-wise probabilistic all-to-all structural connectivity matrix.
- Step 2: Correlation analysis of spatially registered and equally resolved resting activity or multi-stimulus/multi-task activity data (fMRI and/or

MEG) in the same person resulting in a voxel-wise all-to-all functional connectivity matrix.
- Step 3: Cluster comparison between the structural and the functional connectivity matrix identifying regions of consistent structure-function relationships.
- Step 4: Comparison of human analyses (step 3) with structural and functional macaque data to identify correspondences (e.g. visuo-motor pathways) and deviations (speech: fasciculus arcuatus?).
- Step 5: Validation of strongest predictions from final connectivity matrix using custom-designed stimuli and transcranial stimulation in combination with behavioural testing and functional imaging.
- Step 6: Population analysis of healthy subjects and spatial registration to standard brain for probabilistic statements about data from steps 1–5.
- Step 7: Comparison of population data on clustered brain regions to histologically identified regions in probabilistic human brain atlas to assess correspondence.
- Step 8: Comparison of population data between healthy subjects and patient groups supposedly suffering from connectional disorder (e.g., white matter stroke, multiple sclerosis, chronic schizophrenia) using same tasks with similar performance measures.

The steps described here are clearly not the end of this project. However, they are a beginning and give an idea of the opportunities and the challenges that lie ahead of us if we pursue this direction. Even before we tackle the task of relating mesoscopic data to this macroscopic framework, there are enough detailed problems (e.g. concerning spatial normalization of data sets within and across individuals, as well as the homology issue that is often ignored when comparing across species) that need to be addressed.

Conclusions

Anatomical connectivity of the brain is the basis on which physiological interactions unfold and generate the brain's complex dynamics. While it is clear that no such dynamics take place in the absence of connectivity, the precise way in which structure constrains function is still a matter of investigation. The more we know about the anatomical structure the better we can specify these constraints and understand how developmental abnormalities or lesions affect brain function. Modern genetic, morphological and imaging techniques are revealing many clues as to the structural basis of brain function, but they have to be appreciated in the context of their limitations to avoid jumping to conclusions. If used appropriately then anatomy still has much to offer in the study of the nervous system.

Acknowledgments

This work was supported by the DFG (KO 1560/6-2) and by a collaborative grant from the J.S. McDonnell Foundation. I am greatly indebted to C. Opfermann-Rüngeler for drawing Figure 1, and to the editors for their encouragement and patience.

References

Bannister AP. (2005) Inter- and intra-laminar connections of pyramidal cells in the neocortex. Neurosci Res. 53, 95–103.

Barone P, Batardiere A, Knoblauch K, Kennedy H. (2000) Laminar distribution of neurons in extrastriate areas projecting to visual areas V1 and V4 correlates with the hierarchical rank and indicates the operation of a distance rule. J Neurosci. 20, 3263–3281.

Boldogkoi Z, Sik A, Denes A, Reichart A, Toldi J, Gerendai I, Kovacs KJ, Palkovits M (2004) Novel tracing paradigms–genetically engineered herpesviruses as tools for mapping functional circuits within the CNS: present status and future prospects. Prog Neurobiol. 72, 417–445.

Bozkurt A, Kamper L, Stephan K, Kötter R. (2001) Organization of amygdaloprefrontal projections in primate cerebral cortex. Neurocomputing 38–40: 1135–1140.

Bozkurt A, Kamper L, Stephan K, Kötter R. (2002) The structural basis of information transfer from medial temporal lobe to prefrontal cortex in the macaque monkey. Neurocomputing 44–46: 753–758.

Braitenberg V, Schüz A (1998) Cortex, Statistic and Geometry of Neuronal Connectivity. Springer, Berlin.

Burns GA, Young MP (2000) Analysis of the connectional organization of neural systems associated with the hippocampus in rats. Philos Trans R Soc Lond B Biol Sci. 355, 55–70.

Denk W, Horstmann H (2004) Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. PLoS Biol. 2, e329.

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1, 1–47.

Ferbert A, Priori A, Rothwell JC, Day BL, Colebatch JG, Marsden CD. (1992) Interhemispheric inhibition of the human motor cortex. J Physiol. 453, 525–546.

Fiala JC (2002) Three-dimensional structure of synapses in the brain and on the web. Presentation at the World Congress on Computational Intelligence, Honolulu. Downloadable from synapses.bu.edu/lab/fiala/wcci2002.pdf

Gerfen CR, Sawchenko PE. (1984) An anterograde neuroanatomical tracing method that shows the detailed morphology of neurons, their axons and terminals: immunohistochemical localization of an axonally transported plant lectin, Phaseolus vulgaris leucoagglutinin (PHA-L). Brain Res. 290, 219–238.

Germuska M, Saha S, Fiala J, Barbas H (2006) Synaptic distinction of laminar-specific prefrontal-temporal pathways in primates. Cereb Cortex. 16, 865–875.

Gong S, Zheng C, Doughty ML, Losos K, Didkovsky N, Schambra UB, Nowak NJ, Joyner A, Leblanc G, Hatten ME, Heintz N. (2003) A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature 425, 917–925.

Gray EG. (1959) Axo-somatic and axo-dendritic synapses of the cerebral cortex: an electron microscope study. J. Anat. 93, 420–433.
Grillner S, Wallen P. (2002) Cellular bases of a vertebrate locomotor system-steering, intersegmental and segmental co-ordination and sensory control. Brain Res Brain Res Rev. 40, 92–106.
Haeusler S, Maass W (2007) A statistical analysis of information-processing properties of lamina-specific cortical microcircuit models. Cereb Cortex 17, 149–162.
Hatten ME, Heintz N. (2005) Large-scale genomic approaches to brain development and circuitry. Annu Rev Neurosci. 28, 89–108.
Honig MG, Hume RI. (1989) Dil and diO: versatile fluorescent dyes for neuronal labelling and pathway tracing. Trends Neurosci. 12, 333–335 + 340–341.
Hoshi E, Tremblay L, Feger J, Carras PL, Strick PL. (2005) The cerebellum communicates with the basal ganglia. Nat Neurosci. 8, 1491–1493.
Jones DG (1975) Synapses and Synaptosomes. Morphological Aspects. Chapman and Hall, London.
Köbbert C, Apps R, Bechmann I, Lanciego JL, Mey J, Thanos S. (2000) Current concepts in neuroanatomical tracing. Prog Neurobiol. 62, 327–351.
Kötter R (2004) Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. Neuroinformatics 2, 127–144.
Kötter R, Stephan KE (2003) Network participation indices: characterizing component roles for information processing in neural networks. Neural Netw. 16, 1261–1275.
Kötter R, Wanke E. (2005) Mapping brains without coordinates. Phil. Trans. R. Soc. Lond. B. 360: 751–766.
Kötter R, Nielsen P, D.-Johnsen J, Sommer F, Northoff G. (2002) Multi-level neuron and network modeling in computational neuroanatomy. In: Ascoli G. (ed.) Computational Neuroanatomy: Principles and Methods. Humana, Totowa NJ, pp. 359–382.
Le Bihan D. (2003) Looking into the functional architecture of the brain with diffusion MRI. Nat Rev Neurosci. 4, 469–80.
McDonald AJ (1992) Neuroanatomical labeling with biocytin – a review. Neuroreport 3, 821–827.
Nauta WJH (1957) Silver impregnation of degenerating axons. In: New Research Techniques of Neuroanatomy (W.F. Windle, ed.), pp. 17–26. C.C. Thomas, Springfield, Ill.
Rockland KS, Andresen J, Cowie RJ, Robinson DL. (1999) Single axon analysis of pulvinocortical connections to several visual areas in the macaque. J Comp Neurol. 406, 221–250.
Saleem KS, Pauls JM, Augath M, Trinath T, Prause BA, Hashikawa T, Logothetis NK (2002) Magnetic resonance imaging of neuronal connections in the macaque monkey. Neuron 34, 685–700.
Scannell JW, Blakemore C, Young MP (1995) Analysis of connectivity in the cat cerebral cortex. J Neurosci 15, 1463–1483.
Scannell JW, Burns GA, Hilgetag CC, O'Neil MA, Young MP (1999) The connectional organization of the cortico-thalamic system of the cat. Cereb Cortex 9, 277–299.
Schubert D, Kötter R, Zilles K, Luhmann HJ, Staiger JF. (2003) Cell type-specific circuits of cortical layer IV spiny neurons. J. Neurosci. 23: 2961–2970.

Sporns O, Tononi G, Kötter R. (2005) The human connectome: a structural description of the human brain. PLoS Comput Biol 1, e42.

Staiger JF, Kötter R, Zilles K, Luhmann HJ. (1999) Connectivity in the somatosensory cortex of the adolescent rat: an in vitro biocytin study. Anat Embryol (Berl). 199, 357–65.

Stephan KE, Kamper L, Bozkurt A, Burns GA, Young MP, Kötter R (2001) Advanced database methodology for the Collation of Connectivity data on the Macaque brain (CoCoMac). Philos Trans R Soc Lond B Biol Sci. 356, 1159–1186.

Sugino K, Hempel CM, Miller MN, Hattox AM, Shapiro P, Wu C, Huang ZJ, Nelson SB (2006) Molecular taxonomy of major neuronal classes in the adult mouse forebrain. Nat Neurosci. 9, 99–107.

Thomson AM, Bannister AP. (2003) Interlaminar connections in the neocortex. Cereb Cortex. 13, 5–14.

Tuch DS, Wisco JJ, Khachaturian MH, Ekstrom LB, Kotter R, Vanduffel W. (2005) Q-ball imaging of macaque white matter architecture. Philos Trans R Soc Lond B Biol Sci. 360, 869–879.

White J, Southgate E, Thomson J, Brenner S. (1986) The structure of the nervous system of the nematode Caenorhabditis elegans. Phil Trans Roy Soc B 314, 1–340.

Wickersham IR, Finke S, Conzelmann KK, Callaway EM (2007) Retrograde neuronal tracing with a deletion-mutant rabies virus. Nat Methods 4, 47–49.

Yoshimura Y, Dantzker JL, Callaway EM. (2005) Excitatory cortical neurons form fine-scale functional networks. Nature 433, 868–873.

Young MP (1993) The organization of neural systems in the primate cerebral cortex. Proc R Soc Lond Biol Sci 252, 13–18.

Zheng M, Friesen WO, Iwasaki T (2007) Systems-level modeling of neuronal circuits for leech swimming. J Comput Neurosci 22, 21–38.

Zhong YM, Yukie M, Rockland KS. (2006) Distinctive morphology of hippocampal CA1 terminations in orbital and medial frontal cortex in macaque monkeys. Exp Brain Res. 169, 549–553.

Part II

Technology of Structural and Functional Brain Imaging

Primer on Electroencephalography for Functional Connectivity

Thomas C Ferree[1] and Paul L Nunez[2]

[1] Department of Radiology and Bioengineering Graduate Group, University of California, San Francisco and Department of Radiology, University of Texas Southwestern Medical Center, Dallas
[2] Brain Physics, LLC and Department of Biomedical Engineering, Tulane University

1 Introduction

Experimental and theoretical studies of functional connectivity in healthy humans requires non-invasive techniques such as electroenchaphalography (EEG), magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI). Among these, EEG and MEG provide the most direct measure of cortical activity with high temporal resolution ($\lesssim$ 1 msec), but with spatial resolution (1–10 cm) limited by the locations of sensors on the scalp. In contrast, functional MRI has low temporal resolution (1–10 sec), but high spatial resolution (1–10 mm). To the extent that functional activity among brain regions in the cortex may be conceptualized as a large-scale brain network with diffuse nodes, fMRI may delineate the anatomy of these networks, perhaps most effectively in identifying major network hubs. Much current effort is aimed at the integration of these technologies and others, for the obvious reason: to provide the most complete view of dynamical brain activity both spatially and temporally. This chapter focuses primarily on EEG, but also makes connections with MEG.

The human brain exhibits interesting and relevant dynamics on all spatial scales, ranging from single neurons to the entire cortex. The spatial resolution of a particular measurement technique selects certain physiological processes over others. Much investigation in animals has focussed on the information capacities of single neurons, using direct measurements from implanted electrodes. Although EEG measurements integrate over the activity of 10–100 millions of neurons, there is ample evidence that relevant information is represented at these large scales. Indeed, interactions between remote brain areas must involve large spatial scales. Furthermore, several techniques have been

developed for improving the spatial resolution of scalp EEG so that dynamic behavior at the scale of roughly 2–3 cm may be estimated.

The physics and physiology of scalp EEG have been described at length elsewhere (Nunez 1981; Nunez 1995; Nunez and Srinivasan 2006). The goal of this chapter is partly to summarize that material, and partly to extend it. Section 2 describes the physiological genesis of EEG and MEG in terms of cellular currents. Section 3 describes the physical basis of EEG and MEG starting from Maxwell's equations. The remaining chapters focus exclusively on EEG. Section 4 shows how a multipole expansion of the electric potential defines the familiar current dipole. Section 5 adds the effects of head tissue inhomogeneities, which strongly affect the electric potential measured at the scalp. Section 6 reviews EEG measurement principles. Section 7 develops lead field theory, an intuitive way of thinking about the sensitivity of scalp electrodes to brain current sources. Together these sections link concepts of neural activity from the cellular level to the scalp, and provide a basis for the application of scalp EEG to study functional connectivity.

2 Biological Basis of EEG

2.1 Cortical Anatomy

The mammalian cortex is the outer mantle of cells surrounding the central structures, e.g., brainstem and thalamus. It is unique to mammals, and is believed to be necessary for most higher-level brain functions. Topologically the cortex is comprised of two spherical shells, corresponding to the two hemispheres. The hemispheres are connected by the corpus callosum. Cortical thickness varies mostly between 2–3 mm in the human, and is folded around the subcortical structures so as to appear wrinkled. Its average surface area is about 2200 cm^2 (Zilles 1990).

It is estimated that there are roughly 10^{11} neurons in the human brain, and 10^{10} of these in the cortex. Of these, approximately 85% are pyramidal cells (Braitenberg and Schuz 1991), whose dendritic trees have a distinctive, elongated geometry that makes possible the generation of extracellular fields at large distances. The remaining 15% may be broadly classified as stellate cells, whose dendritic trees are approximately spherical, and make little or no contribution to distant fields. Of course, both cells types are interconnected to form a single dynamical network, but it is believed that the fields at large distances are dominated by pyramidal cells.

Synaptic connections in the cortex are dense. Each cortical neuron receives 10^4–10^5 synaptic connections, with most inputs coming from distinct neurons. Pyramidal cells make excitatory connections to both cell types. They make intracortical connections over lengths ranging 0.5–3 mm, and cortico-cortical connections over lengths ranging 1–20 cm. Stellate cells make inhibitory connections to both cell types. They make intracortical connections over lengths

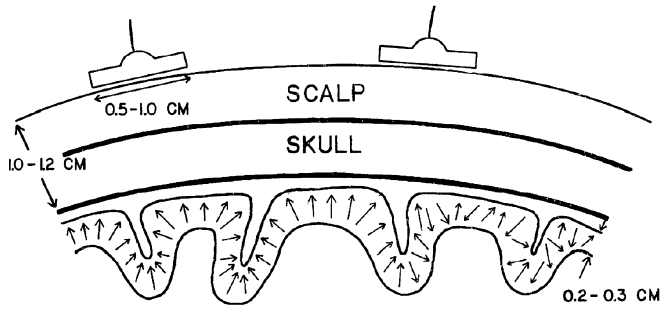


Fig. 1. Depiction of the human head, and the positioning of EEG electrodes relative to the folded cortex. Adapted from Nunez (1995).

ranging only 0.02–0.03 mm. Thus connections in the cortex are said to exhibit long-range excitation and short-range inhibition. Because of the diversity of scales of these synaptic connections, and the nonlocal nature of the cortico-cortical connections, we expect the cortex to exhibit rich spatio-temporal dynamics spanning a wide range of length and time scales.

Figure 1 shows a depiction of several layers of the human head and the positioning of EEG electrodes relative to the cortex. The folds of the cortex are such that even nearby patches of cortex can have different orientations and distances from the detectors. Section 6 shows how individual EEG and MEG detectors spatially integrate neural activity over as much as 100 cm^2. Combining the data from many scalp probes, however, yields an improvement to the order of several cm^2. Using the latter estimate, we must still conclude that EEG and MEG detectors integrate brain activity over a volume including as many as 10^7–10^9 cortical neurons.

2.2 Neurons and Synapses

Neurons are highly specialized for signal processing and conduction via electrochemical processes. The morphological structure of a neuron includes a cell body, called the soma, and elaborate branching structures that enable communication with sensory receptors, distant neurons, etc. In the simplest view, input and output are handled separately. Inputs are collected in a continuous fashion by the dendrites, and represented as a variation of the transmembrane voltage. Multiple inputs are summed in the dendritic tree, and the net input is represented as transmembrane voltage at the soma. When the soma voltage reaches some threshold, a discrete voltage pulse is generated, called an action potential, which propagates down the axon as output. The end of the axon also has elaborate branching to enable communication with target neurons. The input-output properties of neurons have been studied extensively and modeled in detail (Koch and Segev 1989).

2.3 Neuronal Currents

In biological tissues, there are no free electrons. Electric currents are due to ions, e.g., K^+, Na^+, Cl^-, Ca^{2+}, etc. These ions flow in response to the local electric field, according to Ohm's law, but also in response to their local concentration gradient, according to Fick's law (Plonsey 1969). In the resting state of the membrane, the concentration gradients and electric field are due to ion channel pumps, which use energy acquired from ATP to move ions across the membrane against their diffusion gradient.

The concentration of each ion inside and outside the membrane remains essentially constant in time. The transmembrane voltage, however, changes radically in time, the strongest example being the action potential. Thus for the purposes of discussing small scale neural activity, we take the transmembrane potential V_m as the primary dynamic state variable to be considered. By convention, V_m is defined as the potential inside relative to that outside, i.e., $V_\mathrm{m} = V_\mathrm{i} - V_\mathrm{o}$. The current per unit length i_m flowing across the membrane, rather than the transmembrane potential, is considered the basic source variable of the extracellular fields detected by EEG and MEG.

Currents flow in neurons when a neurotransmitter binds to receptors on ion channels located in the membrane. Detailed consideration of a three-dimensional dendrite or axon has shown that the phenomenon of current flow in the cell may be well described by a one-dimensional approximation. The same mathematical treatment applies to undersea telegraph cables, comprised of an insulated metal core immersed in conductive salt water, thus the treatment is called "cable theory."

Assuming a membrane with conductive and capacitive properties, surrounded by fluids with conductive properties only, and applying current conservation in each compartment leads to

$$\tau_\mathrm{m} \frac{\partial V_\mathrm{m}}{\partial t} = \lambda_\mathrm{m}^2 \frac{\partial^2 V_\mathrm{m}}{\partial x^2} - (V_\mathrm{m} - E_\mathrm{r}) - r_\mathrm{m}\, i_\mathrm{other}(x,t) \tag{2.1}$$

where the membrane time constant $\tau_\mathrm{m} = r_\mathrm{m} c_\mathrm{m}$, the neuron space constant $\lambda_\mathrm{m} = \sqrt{r_\mathrm{m}/(r_\mathrm{i} + r_\mathrm{o})}$, the membrane resistance times unit area $r_\mathrm{m} = 1/g_\mathrm{m}$, and E_r is the resting transmembrane potential. Mathematically this is identical in form to the heat equation, which governs the one-dimensional flow of heat in a heat-conducting rod. It has been studied extensively and has well-known analytic solutions for appropriate boundary conditions.

Consider an isolated dendrite of length L, with current injected at one end. A solution may be derived for $V_m(x)$. The corresponding transmembrane current per unit length $i_m(x)$ may be written

$$i_\mathrm{m}(x) \simeq -I_0\, \delta(x) + \sqrt{\frac{r_\mathrm{i}}{r_\mathrm{m}}} I_\mathrm{inj}\, \mathrm{e}^{-x/\lambda_\mathrm{m}}, \qquad x \geq 0 \tag{2.2}$$

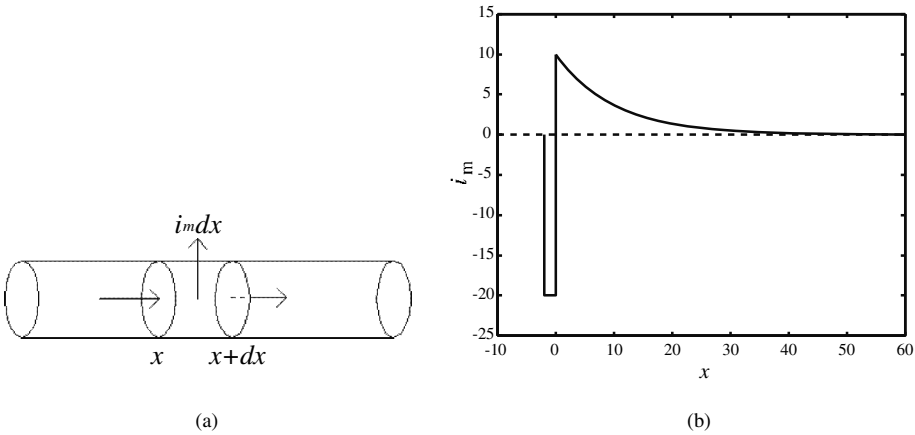


Fig. 2. (**a**) Cylindrical cable representing a segment of dendrite or axon. A single compartment is shown with transmembrane current per unit length i_m. (**b**) Transmembrane current per unit length i_m at steady-state, due to current injected at $x = 0$. (The delta function at $x = 0$ is not drawn to scale; the actual area of the delta function equals the area of the positive (outward) current flow.)

where $\delta(x)$ is the Dirac delta function. Figure 2(b) shows the solution for $x \geq 0$. The delta function in Fig. 2(b) is not drawn to scale; the integral of (2.2) over all x is zero. Section 4 shows that, far away from the cable, the potential Φ may be approximated as if it were generated by an ideal dipole, consisting of a point source and sink separated by a distance λ_m. This transmembrane current, driven not so much by the transmembrane potential as by difffusion, implies a nonzero extracellular electric field through current conservation.

2.4 Large Neural Populations

Neuronal currents of this sort generate extracellular electric and magnetic fields, which are detected using EEG and MEG. The fields generated by a single neuron are much too small to be detected at the scalp, but the fields generated by synchronously active neurons, with advantageous geometric alignment, can be detected. Stellate cells have approximately spherical dendritic trees, so far away the extracellular fields tend to add with all possible orientiations and cancel. Pyramidal cells have similar dendritic trees, but the tree branches are connected to the cell body (soma) by a long trunk, called the apical dendrite. It is a fortuitous anatomical feature of the cortex that pyramidal cells have their apical dendrites aligned systematically, along the local normal to the cortical surface. In this way, the fields of pyramidal neurons superimpose geometrically to be measurable at the scalp.

Several factors contribute to the net fields measured at the scalp. Many active neurons and fortuitous alignment are not enough. As neuronal oscillations tend to oscillate with predominant frequencies, dependent upon functional activity, only synchronously active neurons will sum coherently in time (Elul 1972; Nunez 1981). Consider a 1 cm^2 region of cortex, containing approximately 10^7 aligned pyramidal cells. Make the idealized assumption that all these neurons are oscillating at the predominant frequency, (e.g., 10 Hz resting rhythm). If only 1% of these neurons are synchronously active, i.e., oscillate in phase with each other, and 99% are oscillating with random phase. If the contribution from the asynchronous neurons may be treated as Gaussian then, because N unit-variance Gaussian random numbers sum to $\sqrt{N}$, the relative contribution of synchronous to asynchronous neurons would be $10^5/\sqrt{10^7} \simeq 30$. Thus scalp EEG and MEG are considered to be dominated by synchronous neural activity. Indeed, amplitude reduction in EEG clinical and research circles is often termed desynchronization. Of course, phase desynchronization is only one of many possible mechanisms that could reduce the amplitude of the net voltage at the scalp. Alternatively, synchronous spike input to a patch of cortex can generate event-related synchronization in the dendritic fields. Such phase synchronization is one mechanism for producing event-related potentials (Makeig et al. 2002).

3 Physical Basis of EEG

3.1 Electromagnetic Fields in Conductive Media

The physics of electric and magnetic fields in matter is summarized by Maxwell's equations. This set of coupled, linear equations has source terms given by the charge density ρ and the current density $\mathbf{J}$. Additional contributions arise from the time derivatives of the fields. In matter the macroscopic fields obey (Jackson 1975)

$$\vec{\nabla} \cdot \mathbf{D} = \rho \tag{3.1}$$

$$\vec{\nabla} \cdot \mathbf{B} = 0 \tag{3.2}$$

$$\vec{\nabla} \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{3.3}$$

$$\vec{\nabla} \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{3.4}$$

where $\mathbf{E}$ is called the electric field, and H is called the magnetic field[3]. The electric displacement $\mathbf{D}$ is related to the electric field $\mathbf{E}$ through ϵ the dielectric constant: $\mathbf{D} = \epsilon\mathbf{E}$. The magnetic induction $\mathbf{B}$ is related to the electric field $\mathbf{H}$ through μ the magnetic susceptibility: $\mathbf{B} = \mu\mathbf{H}$.

[3] Here Maxwell's equations are expressed in MKS units. The equations appear different from those in CGS units (Jackson 1975), but MKS is the more common in bioelectromagnetism (Gulrajani 1998).

Maxwell's equations reflect the basic principle of charge conservation. Taking the divergence of (3.4) and the time derivative of (3.1) leads to

$$\vec{\nabla} \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0 \tag{3.5}$$

Integrating over a closed volume V bounded by a surface S and using the divergence theorem (Arfken 1995) shows that the component of the current flowing outward across S equals minus the time rate of change of the charge in the volume bounded by S.

3.2 Macroscopic Source Current J_S

Biological tissues have conductive and capacitive properties, but the magnetic susceptibility is essentially that of vacuum. In metals the charge carriers are free electrons, but in biological tissue the charge carriers are ions, e.g., Na^+, K^+, Cl^-, Ca^{2+}, etc. Section 2 described how the membrane current density $\mathbf{J}$ has both electric and diffusive contributions. In the extracellular space, the story is more complicated. There is no concentration gradient or diffusive contribution *per se*, nevertheless, the fields in the extracellular space may be computed by considering the current density to have two contributions:

$$\begin{aligned} \mathbf{J} &= \mathbf{J}_\mathrm{E} + \mathbf{J}_\mathrm{S} \\ &= \sigma \mathbf{E} + \mathbf{J}_\mathrm{S} \end{aligned} \tag{3.6}$$

where $\mathbf{J}_\mathrm{E}$ is the *ohmic* current that flows in response to the local electric field, and $\mathbf{J}_\mathrm{S}$ is the *source* current (or *impressed* current). Within the membrane, (2.1) includes contributions to the current arising from both the transmembrane electric field and the transmembrane concentration gradients for each ion species. For computing fields in the extracellular space, $\mathbf{J}_\mathrm{S}$ is a phenomenological device that subsumes other physical aspects of the problem (Plonsey 1982; Nunez and Srinivasan 2006).

3.3 Solution to Maxwell's Equations

Maxwell's equations may be solved analytically if the parameters ϵ, μ, and σ are constant in space. This ideal case forms the basis of solutions in systems with piecewise constant parameters. The solution to (3.1–3.4) for the fields $\mathbf{E}$ and $\mathbf{B}$ is obtained by introducing the magnetic (vector) potential $\mathbf{A}$, defined by

$$\mathbf{B} = \vec{\nabla} \times \mathbf{A} \tag{3.7}$$

and the electric (scalar) potential Φ, defined by

$$\mathbf{E} = -\vec{\nabla}\Phi - \frac{\partial \mathbf{A}}{\partial t} \tag{3.8}$$

Because these equations involve the curl of $\mathbf{A}$ and the divergence of Φ, and there are vector identities specifying the conditions in which the divergence and curl vanish, there is additional flexibility in defining these potentials. This flexibility is called gauge invariance, and by choosing a convenient gauge:

$$\vec{\nabla} \cdot \mathbf{A} + \mu\epsilon \frac{\partial \Phi}{\partial t} + \mu\sigma\Phi = 0 \tag{3.9}$$

the differential equations for $\mathbf{A}$ and Φ separate (Gulrajani 1998).

Assuming harmonic time dependence $\Phi(\mathbf{r}, t) = \mathrm{Re}[\tilde{\Phi}(\omega, t)\mathrm{e}^{i\omega t}]$, the uncoupled equations have the well-known solutions (Arfken 1995).

$$\tilde{\Phi}(\mathbf{r}, \omega) = \frac{-1}{4\pi(\sigma + i\omega\epsilon)} \int_{\mathrm{V}} \frac{\vec{\nabla}' \cdot \tilde{\mathbf{J}}_{\mathrm{S}}(\mathbf{r}', \omega)}{|\mathbf{r} - \mathbf{r}'|} \, \mathrm{e}^{-ik|\mathbf{r}-\mathbf{r}'|} \, d^3r' \tag{3.10}$$

and

$$\tilde{\mathbf{A}}(\mathbf{r}, \omega) = \frac{\mu}{4\pi} \int_{\mathrm{V}} \frac{\tilde{\mathbf{J}}_{\mathrm{S}}(\mathbf{r}', \omega)}{|\mathbf{r} - \mathbf{r}'|} \, \mathrm{e}^{-ik|\mathbf{r}-\mathbf{r}'|} \, d^3r' \tag{3.11}$$

These solutions are valid at any frequency, and are therefore useful in electrical impedance tomography (EIT) and transcranial magnetic stimulation (TMS), where the electric and magnetic fields are controlled by an external device that may be driven to high frequencies, e.g., ∼100kHz. When applied to EEG and MEG, however, where the frequencies are limited physiologically, these equations may be simplified by the approximation $\omega \to 0$. This is called the quasi-static limit: the fields at each time point t are computed from the sources at that same time point, with no electromagnetic coupling or propagation delays related to the speed of light.

3.4 Quasistatic Formulation

Electric Potential

The solutions for Φ and $\mathbf{A}$ may be derived more directly by taking the quasistatic limit at the start, i.e., setting all time derivatives in Maxwell's equations equal to zero. The differential statement of current conservation (3.5) becomes

$$0 = \vec{\nabla} \cdot \mathbf{J} = \vec{\nabla} \cdot \mathbf{J}_{\mathrm{E}} + \vec{\nabla} \cdot \mathbf{J}_{\mathrm{S}} \tag{3.12}$$

Substituting $\mathbf{E} = -\vec{\nabla}\Phi$ leads to

$$\vec{\nabla} \cdot (\sigma\vec{\nabla}\Phi) = \vec{\nabla} \cdot \mathbf{J}_{\mathrm{S}} \tag{3.13}$$

Assuming σ is constant reduces this to Poisson's equation

$$\nabla^2\Phi = \frac{1}{\sigma}\vec{\nabla} \cdot \mathbf{J}_{\mathrm{S}} \tag{3.14}$$

which has the well-known solution

$$\Phi(\mathbf{r},t) = \frac{-1}{4\pi\sigma}\int_{\mathrm{V}} \frac{\vec{\nabla}'\cdot\mathbf{J}_{\mathrm{S}}(\mathbf{r}',t)}{|\mathbf{r}-\mathbf{r}'|}\, d^3r' \tag{3.15}$$

Thus the electric potential Φ may be computed at each time point t as though $\mathbf{J}_{\mathrm{S}}$ were constant in time.

Magnetic Field

Similarly, for the magnetic field,

$$\nabla^2\mathbf{B} = -\mu_0\vec{\nabla}\times\mathbf{J} \tag{3.16}$$

Assuming σ is constant reduces this to a simpler expression in terms of the source current $\mathbf{J}_{\mathrm{S}}$ only

$$\nabla^2\mathbf{B} = -\mu_0\vec{\nabla}\times\mathbf{J}_{\mathrm{S}} \tag{3.17}$$

This is Poisson's equation for each component of $\mathbf{B}$, and has the solution

$$\mathbf{B}(\mathbf{r},t) = \frac{\mu_0}{4\pi}\int_{\mathrm{V}} \frac{\vec{\nabla}'\times\mathbf{J}_{\mathrm{S}}(\mathbf{r}',t)}{|\mathbf{r}-\mathbf{r}'|}\, d^3r' \tag{3.18}$$

The fundamental similarity between electric and magnetic fields, even when uncoupled at low frequencies, leads to parallel expressions of the basic EEG and MEG equations.

4 Dipole Source Modeling

4.1 Multipole Expansion of Φ

Equations (3.15) and (3.18) are the general solutions for Φ and $\mathbf{B}$ given an arbitrary current source density $\mathbf{J}_{\mathrm{S}}$ in the absence of boundaries. The integrand of each function involves derivatives of $\mathbf{J}_{\mathrm{S}}(\mathbf{r}')$, and the integration kernel $1/|\mathbf{r}-\mathbf{r}'|$ called the Green's function (Jackson 1975). The basis of the multipole expansion is to assume that $\mathbf{J}_{\mathrm{S}}(\mathbf{r}')$ is confined to some finite region of space, and that the point $\mathbf{r}$ at which the field is being computed or measured is far away compared to the size of the source distribution, i.e., $|\mathbf{r}| \gg |\mathbf{r}'|$. Computing the Taylor series of $1/|\mathbf{r}-\mathbf{r}'|$ through the first two terms gives

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|} = \frac{1}{|\mathbf{r}|} + \frac{\mathbf{r}\cdot\mathbf{r}'}{|\mathbf{r}|^3} + \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\frac{3r_ir_j-\delta_{ij}|\mathbf{r}|^2}{|\mathbf{r}|^5}r_i'r_j' + \ldots, \quad |\mathbf{r}'| \ll |\mathbf{r}| \tag{4.1}$$

The first term is called the *monopole* term, and falls off as $1/|\mathbf{r}|$. The second term is called the *dipole* term, and falls off as $1/|\mathbf{r}|^2$. The third term is called

the quadrupole term, and falls off as $1/|\mathbf{r}|^3$, and so on. Inserting (4.1) into (3.15) gives

$$\Phi(\mathbf{r}) = \Phi^{(1)}(\mathbf{r}) + \Phi^{(2)}(\mathbf{r}) + \Phi^{(3)}(\mathbf{r}) + \dots \tag{4.2}$$

and so on.

The monopole term is

$$\begin{aligned}\Phi^{(1)}(\mathbf{r}) &= \frac{-1}{4\pi\sigma}\frac{1}{|\mathbf{r}|}\int_V \vec{\nabla}' \cdot \mathbf{J}_S(\mathbf{r}')\; d^3r' \\ &= \frac{-1}{4\pi\sigma}\frac{1}{|\mathbf{r}|}\int_S \mathbf{J}_S(\mathbf{r}') \cdot \hat{\mathbf{n}}\; dS'\end{aligned} \tag{4.3}$$

where the second equality follows from the divergence theorem. If the volume V contains as many source as sinks (of equal strength), then the monopole term vanishes by current conservation.

The dipole term is

$$\begin{aligned}\Phi^{(2)}(\mathbf{r}) &= \frac{1}{4\pi\sigma}\left[\int_V \mathbf{r}'\left(\vec{\nabla}' \cdot \mathbf{J}_S(\mathbf{r}')\right) d^3r'\right] \cdot \vec{\nabla}\frac{1}{|\mathbf{r}|} \\ &= \frac{1}{4\pi\sigma}\; \mathbf{p} \cdot \vec{\nabla}\frac{1}{|\mathbf{r}|}\end{aligned} \tag{4.4}$$

where the *dipole moment* is defined as

$$\mathbf{p} = \int \mathbf{r}'\left(\vec{\nabla}' \cdot \mathbf{J}_S(\mathbf{r}')\right) d^3r' \tag{4.5}$$

Because the monopole term normally vanishes, and the quadrupole term falls off more quickly with distance, the dipole term usually makes the largest contribution to Φ and $\mathbf{B}$.

4.2 Electric Dipoles

Point Sources

A current dipole may be idealized as a source and sink with equal magnitude, separated by an infinitesimal distance d. This may be written formally as

$$\vec{\nabla} \cdot \mathbf{J}_S = -I_0 \lim_{d\to 0}\left[\delta^{(3)}(\mathbf{r} - \mathbf{r}_+) - \delta^{(3)}(\mathbf{r} - \mathbf{r}_-)\right] \tag{4.6}$$

where $\mathbf{r}_+$ ($\mathbf{r}_-$) is the source (sink) location, and $\mathbf{d} \equiv \mathbf{r}_- - \mathbf{r}_+$ is the directed distance from sink to source.[4] Inserting (4.6) into (4.5) leads to an intuitive expression for the dipole moment:

$$\mathbf{p} = I_0\mathbf{d} \tag{4.7}$$

where I_0 is the magnitude of the current, and $\mathbf{d}$ is the directed distance from source to sink.

[4] Technically, the limit $d \to 0$ refers to the dual limit: $d \to 0$ and $I_0 \to \infty$, such that the product $p = I_0 d$ remains constant and finite.

Dendritic Cable

The current distribution shown in Fig. 2(b) may be written

$$\vec{\nabla} \cdot \mathbf{J}_{\mathrm{S}} = +I_0 \ \delta(x)\delta(y)\delta(z) - i_{\mathrm{m}}(x)\delta(y)\delta(z) \tag{4.8}$$

In both terms, the factors $\delta(y)\delta(z)$ ensure that the source lies on the x-axis. In the first term, the factor $\delta(x)$ puts the sink at $x = 0$. In the second term, the transmembrane current per unit length $i_{\mathrm{m}}(x)$ is given by (2.2).

Inserting (4.8) into (4.3) gives

$$\int \vec{\nabla} \cdot \mathbf{J}_{\mathrm{S}}(\mathbf{r}) \ d^3r = \int_0^\infty [I_0 \ \delta(x) - i_{\mathrm{m}}(x)] \ dx = 0 \tag{4.9}$$

where the last equality follows from direct integration of (2.2). Thus the monopole contribution vanishes by current conservation, i.e., the total of sources and sinks equals zero.

Inserting (4.8) into (4.5) gives

$$\mathbf{p} = \int \mathbf{r} \, [I_0 \ \delta(x) - i_{\mathrm{m}}(x)] \, \delta(y)\delta(z) \ d^3r \tag{4.10}$$

The three vector components may be evaluated separately. Because of the factor $\mathbf{r} = (x, y, z)$, integration over y and z gives $p_y = 0$ and $p_z = 0$, respectively. Similarly for p_x, integrating over x causes the first term involving $\delta(x)$ to vanish, leaving

$$p_x = -\int_0^\infty x \ i_{\mathrm{m}}(x) \ dx = -I_0 \ \lambda_{\mathrm{m}} \tag{4.11}$$

This result is intuitive: For the ideal dipole (4.6), the dipole moment (4.7) is equal to the current I_0 times the distance d between the source and sink. For the uniform cable with current injected at one end (4.11), the distance d is replaced by λ_{m}, the characteristic length scale for the decay of the current along the dendrite. The minus sign indicates that the direction of extracellular current flow is leftward in Fig. 2(b).

Extracellular Fields of Axons

There are three main arguments that axons make negligible contribution to EEG-MEG. First, the quadrupolar field falls of rapidly with distance, and is likely to be dominated at the scalp by the dipolar fields of dendrites. Second, axons are not arranged systematically in the cortex as are the apical dendrites of pyramidal cells, thus the geometric superposition of axonal fields can not occur to the same degree as it does for dendrites. Third, action potentials have ~ 1 ms duration, and therefore have a dominant frequency near 1000 Hz. The EEG-MEG signal has most of its power below 100 Hz, which is more like the

time scale over which dendritic potentials vary. The extracellular field due to a single neuron is not detectable at the scalp, but the superposition of many synchronously active neurons is detectable. In order for action potential fields to superimpose to measurable levels at the scalp, it would be necessary for the action potentials of multiple neurons to occur with high temporal synchrony. While it does appear that spike synchrony plays a fundamental role in neural processing, the requirements on synchrony are much more demanding for axons than dendrites due to the shorter duration of their potentials. For these reasons, it is expected that EEG-MEG is dominated by dendritic potentials. Up to the size of the integration volume of an EEG or MEG sensor, dipolar sheets created by synchronously active patches of cortex make contributions to the scalp potential in proportion to their size.

For each argument against axon contributions, there is a reasonable counter-argument, First, dipole fields likely dominate the scalp potential, but that does not mean that quadrupole fields are totally negligible. Second, axons run in fiber bundles, and synchronous input to their neurons generates a compound action potential. Third, sensory input typically generates a neural response with abrupt onset and high degree of neural synchrony, at least at the dendritic level. This increases the firing probability in time and can increase spike synchrony. Thus spike synchrony in fiber bundles could potentially superimpose to be measurable at the scalp. Thus, although cortical pyramidal dendrites likely dominate axonal fields in resting EEG, action potentials could conceivably contribute to the scalp potential, particularly in early sensory response. Still, the short duration of spikes puts their power at high frequencies, which are filtered out in many EEG recordings.

5 Human Head Modeling

So far we have assumed an infinite conducting medium, ignoring tissue boundaries and head geometry. The conductivity σ changes abruptly at tissue boundaries, and has major effects on EEG signals recorded at the scalp. Effects on MEG are smaller, but nonzero. The "forward" problem of EEG: Given the brain current sources, compute the electric potential Φ on the scalp are the locations of the measurement and reference electrodes. Since Φ depends linearly on the source currents $\mathbf{J}_S$, linear superposition applies, and it is sufficient to compute the fields due to a single dipole at first.

5.1 Mathematical Considerations

Boundary Conditions

Consider a four-layer head model with conductivity parameters σ_a, where $a = 1, \ldots, 4$, and define the conductivity of air to be $\sigma_5 = 0$. Let Φ_a be the potential in layer a, and let $\hat{\mathbf{n}}$ be the outward-oriented normal to the

boundary surface S. In each layer, σ is constant and the homogenous solution is correct. The solutions to Poisson's equation in each layer are joined together by appropriate boundary conditions.

The first condition boundary condition is that the normal current density $\mathbf{J}_\perp$ be continuous across each boundary:

$$\sigma_a \frac{\partial \Phi_a}{\partial n}\bigg|_S = \sigma_{a+1} \frac{\partial \Phi_{a+1}}{\partial n}\bigg|_S \tag{5.1}$$

where the normal derivative is defined $\partial\Phi/\partial n \equiv \vec{\nabla}\Phi_a \cdot \hat{\mathbf{n}}$. From Maxwell's equation a second boundary condition may be shown: continuity of the parallel component of the electric field. Assuming no sources or sinks on the surface, this is equivalent to continuity of the potential Φ across each boundary:

$$\Phi_a\Big|_S = \Phi_{a+1}\Big|_S \tag{5.2}$$

as may be shown by drawing a rectangular loop with one side in each layer, and integrating the electric field around this loop.

The magnetic field $\mathbf{B}$ obeys similar boundary conditions involving discontinuities in μ (Jackson 1975). These are not relevant to biological tissue, because to high accuracy $\mu = \mu_0$, the magnetic susceptibility of vacuum (Plonsey 1969). Never must we consider discontinuities in μ or boundary conditions on $\mathbf{B}$ in the usual sense. Yet boundary effects do enter at tissue discontinuities. In passing from (3.16) to (3.17) we assumed σ to be constant. Without that assumption we have additional contributions to $\mathbf{B}$ arising from discontinuities in σ. These contributions are identically zero for a spherical head model, but nonzero in general.

Uniqueness of Solution

Poisson's equation for Φ has a unique solution given an appropriate specification of the boundary conditions across the entire surface, including: 1) the potential Φ, or 2) its normal derivative $\partial\Phi/\partial n$, is specified on the boundary (Jackson 1975). These are called Dirichlet and Neumann boundary conditions, respectively. Mixed boundary conditions are also possible, in which Φ and $\partial\Phi/\partial n$ are are specified on non-overlapping parts of the boundary. (Specifying both Φ and $\partial\Phi/\partial n$ over any part of the boundary is an overspecification of the problem, and the existence of a solution is not guaranteed.) This uniqueness property allows us to be creative in how we derive the solution, since finding *any* solution to Poisson's equation which matches the boundary conditions implies that we have found *the* solution.

5.2 Spherical Head Method

The simplest head model that accommodates the layered tissues is comprised an inner sphere (brain) surrounded by 2 (ignoring CSF) or 3 (including CSF)

concentric spheres (see Fig. 1). For a dipole current source at brain location **a** with dipole moment **m**, the potential at the scalp surface location **r** may be written compactly

$$\Phi(\mathbf{r}) = \sum_{n=1}^{\infty} c_n f^{n-1} \mathbf{m} \cdot \left[\hat{\mathbf{r}}\ P_n(\cos\theta) + \hat{\mathbf{t}}\ \frac{P_n^1(\cos\theta)}{n} \right] \tag{5.3}$$

where $f \equiv a/r_4$ is the dipole eccentricity, r_4 is the outer scalp radius, θ is the angle between $\hat{\mathbf{r}}$ and $\hat{a}$, $\hat{\mathbf{r}}$ is the radial unit vector, $\hat{\mathbf{t}}$ is the tangential unit vector, and the c_n are constant coefficients (Salu et al. 1990). Current conservation ensures that the surface integral of the absolute potential Φ induced by a dipolar current source is zero. This is reflected in (5.3) by the absence of a constant term that would be represented by $n = 0$. Thus the potential Φ computed with (5.3) is implicitly referenced to infinity.

In numerical implementations of (5.3), the calculation of the Legendre polynomials $P_n(x)$ is the rate limiting step. Faster implementation is available by noting the convergence properties of the series (Sun 1997).

5.3 Boundary Element Method

The simplest approach for accommodating realistic head geometry keeps the assumption that the head is comprised of four tissue layers: brain, CSF, skull and scalp, and that each layer is described by a single homogeneous and isotropic conductivity σ, but relaxes the assumption of sphericity. Green's theorem may be used to write the solution to Poisson's equation as an integral equation for Φ

$$\Phi(\mathbf{r}_o) = \frac{2\sigma_1}{\sigma_o + \sigma_{o+1}} \Phi_\infty(\mathbf{r}_o) + \frac{1}{2\pi} \sum_{a=1}^{4} \frac{\sigma_a - \sigma_{a+1}}{\sigma_a + \sigma_{a+1}} \int_{\Gamma_a} \Phi(\mathbf{r})\ d\Omega_{\mathbf{r}_o\mathbf{r}} \tag{5.4}$$

where

$$d\Omega_{\mathbf{r}_o\mathbf{r}} \equiv \frac{(\mathbf{r} - \mathbf{r}_o)}{|\mathbf{r} - \mathbf{r}_o|^3} \cdot \hat{\mathbf{n}}\ \mathrm{dS} \tag{5.5}$$

is the solid angle subtended by the surface element dS at $\mathbf{r}$, as viewed from the observation point $\mathbf{r}_o$ (Barnard et al. 1967a; Barnard et al. 1967b; Geselowitz 1967). This equation shows how Φ at each point $\mathbf{r}_o$ in V depends upon the integral of Φ over each tissue boundary surface S, and that the surface contributions are of the same form as a surface dipole layer.

In numerical implementations of (5.4), the basic approach is to discretize the surface with a set of triangular elements, and evaluate the surface integral as a discrete sum. Figure 3 shows surface meshes for this purpose. In setting up the sum, the potential on the surface may be expressed in terms of either the potentials at the corners, or the potentials on the faces. The former is faster computationally because the number of corners is approximately half the number of faces (Barr et al. 1977). It also allows an improvement in which

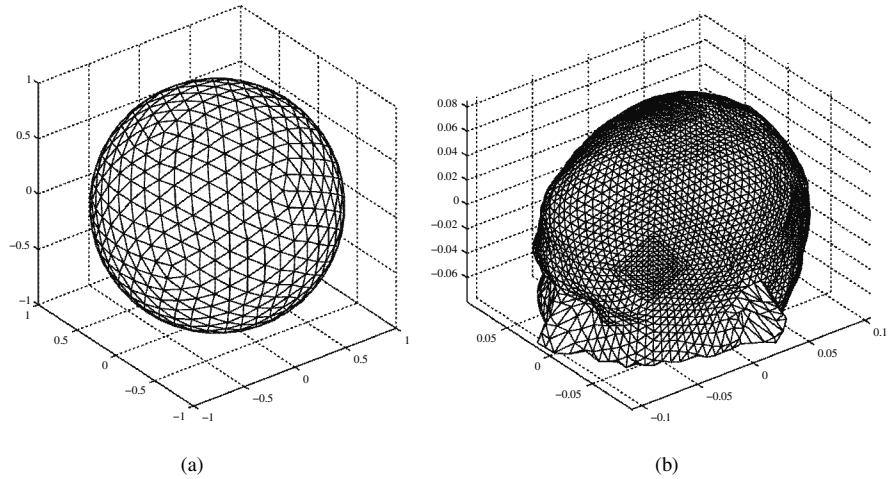


Fig. 3. (**a**) Surface mesh for (**a**) sphere with 1280 faces, (**b**) human scalp with 8192 faces. Meshes are generated on a spherical surface, then fit to the scalp-air boundary of a high-resolution structural MRI

the potential varies linearly over each triangle (Gençer et al. 1999). Evaluating $\mathbf{r}_o$ at each corner leads to a matrix equation for Φ at the corners, which may be solved by inversion. Once Φ is known on S, then (5.4) may be evaluated at any $\mathbf{r}_o$ in V. Scalp potential values at the electrodes may be computed this way, or estimated using spline interpolation (see Sect. 8).

5.4 Conductive Properties of Head Tissues

Aside from inhomogeneities and anisotropies ignored in spherical head models, the conductivity of head tissues are known within some (perhaps large) range of error (Foster and Schwan 1989). The brain conductivity $\sigma_1 \simeq 0.15$ S/m (Stoy et al. 1982). The CSF conductivity $\sigma_2 \simeq 1.79$ S/m (Baumann et al. 1997). The scalp conductivity $\sigma_4 \simeq 0.44$ S/m (Geddes and Baker 1967). The conductivity of the living human skull, however, has been a source of mass confusion. Rush and Blanchard (1966) measured the conductivity ratio between the skull and that of saline in which the skull was immersed, and found conductivity ratios ranging from 50 to 300. Rush and Driscoll (1968) found a ratio near 80, then applied that ratio between the brain and skull, as though the living skull were saturated with brain-like rather than saline-like fluid. Most subsequent studies (e.g., Stok 1987) have used this ratio. Assuming the brain conductivity $\sigma_1 \simeq 0.15$ S/m, for example, $\sigma_1/\sigma_3 \simeq 80$ implies $\sigma_3 \simeq 0.002$ S/m.

Since then evidence has accumulated that this early reasoning may greatly underestimate σ_3. Even within the context of the Rush and Driscoll (1968) study, assuming the saline conductivity $\sigma \simeq 1.3$ S/m implies $\sigma_3 \simeq 0.017$ S/m. Kosterick et al. 1984 reported $\sigma_3 \simeq 0.012$ S/m. Averaging the values reported in Law et al. (1993) suggests $\sigma_3 \simeq 0.018$ S/m. Oostendorp et al.

(2000) reported $\sigma_3 \simeq 0.015$ S/m. This series of literature seems to implies consistently that $\sigma_3 \simeq 0.015$ S/m and $\sigma_1/\sigma_3 \simeq 10$. This ratio is lower than the range 20–80 suggested by Nunez and Srinivasan (2005), due partly to a lower estimate of brain conductivity. With this skull conductivity, assuming the brain conductivity $\sigma_1 \simeq 0.33$ S/m (Stok 1987), for example, gives the ratio $\sigma_1/\sigma_3 \simeq 22$. Early models assumed the brain and scalp conductivity were equal (Rush and Driscoll 1968). If this skull conductivity is compared to the scalp rather than the brain, $\sigma_4/\sigma_3 \simeq 29$.

As discussed in Nunez and Srinivasan (2006), however, the effective conductivity of a single layered skull used in a volume conductor model may be substantially lower than its actual conductivity due to several shunting tissues not included in such models, e.g., extra CSF, the middle skull layer, and the anisotropic white matter. For example, consider a three-layered skull in which the inner skull layer conductivity is substantially higher than the inner and outer skull layers (as verified experimentally). Imagine a limiting case where the resistivity of the inner layer goes to zero so that no current enters the outer skull layer or scalp (zero scalp potential everywhere). The effective brain-to-skull conductivity ratio is infinite in this limiting case, even though the measured value based on a composite skull could easily be less than 20. This argument implies that effective brain-to-skull conductivity ratios cannot be accurately estimated from impedance measurements of composite skull tissue alone.

6 Data Acquisition

6.1 Electrode and Amplifier Systems

In EEG recordings, electric potential is measured on the scalp surface, and used to make inferences about brain activity. Although potentials relative to infinity are often considered in theoretical derivations, in practice only potential differences can be measured. Thus EEG measurements always involve the potential difference between two sites. This is accomplished using differential amplifiers, which include the measure electrodes, a reference electrode, and an "isolated common" electrode that takes the place of true ground.

Huhta and Webster (1973) presented an essentially complete analysis of electrocardiographic (ECG) recordings using differential amplifiers, including signal loss and 60 Hz noise. Several of their assumptions are either outdated or not applicable to EEG. First, they assumed that the subject was resistively coupled to earth ground. This simplification reduces the number of variables in the calculations, but is unsafe because it increases the risk of electric shock. It also allows more 60 Hz noise to enter the measurements because ambient electric potential fields in the recording environment exist relative to earth ground. Second, they assumed the grounding electrode was connected to the subjects foot, at maximal distance from the recording and reference electrodes

which were located on the torso for cardiac recording. The thinking was that the foot would be electrically quiet, which may be true, but this increases 60 Hz noise because the entire body acts as an antenna.

Modern EEG systems are designed differently (Ferree et al. 2001). First, safety regulations require that the subject be isolated from ground so that contact with an electric source would not result a path to earth ground. This is accomplished by using an "isolated common" electrode that is electrically isolated from the ground of the power supply. In this configuration, the subject is only capacitively coupled to true ground, largely eliminating the risk of electric shock, and reducing 60 Hz noise. The measurement is then made as follows. The potential of both measurement and reference leads are taken relative to the common electrode, then their difference is amplified. Second, both the reference and common electrodes are located on the head in order to minimize 60 Hz common-mode noise sources, as well as physiological noise from cardiac sources.

6.2 Temporal Sampling

The validity of the quasi-static approximation to Maxwell's equations in biological materials is equivalent to saying that the electric and magnetic fields propagate from the brain to the scalp instantaneously. In this sense, the temporal resolution of EEG (and MEG) is unlimited. Because most of the power in EEG time series falls below 100 Hz, typical sampling rates are 250 Hz, 500 Hz, and 1 kHz. Higher sampling rates are used to measure the brain-stem auditory evoked potential, and to adequately represent artifacts when EEG is recorded simultaneously with fMRI, but usually lower sampling rates are preferred because they result in smaller file sizes and faster analysis.

In digital signal processing, the Nyquist theorem states that power at frequency f in a signal must be sampled with interval $\Delta t \leq 1/(2f)$. For fixed Δt, this means that only frequencies $f \leq 1/(2\Delta t)$ are accurately represented; this is called the Nyquist frequency. Power at frequencies $f > 1/(2\Delta t)$ are aliased, i.e., represented inaccurately as power at lower frequencies. To avoid this, EEG and other amplifiers sample in two stages. For a given choice of sampling rate Δt, analog filters are applied to remove signal power at frequencies $f > 1/(2\Delta t)$, then the signal is sampled discretely. In this way, EEG amplifiers have a wide range of sampling rates that may be selected without aliasing.

6.3 Spatial Sampling

In clinical EEG, speed, convenience, and culture typically dictate that only 19 electrodes be used, with inter-electrode spacing around 30 degrees. This configuration reveals large-scale brain activity reasonably well, and taking the potential difference between neighboring electrode pairs can isolate focal activity between those electrodes provided other conditions are met. Generally speaking, however, this low density misses much of the spatial information in the scalp potential. In research EEG, electrode arrays typically have 32, 64,

128, or 256 recording channels. The more electrodes, the more information, but there is a limit to the improvement.

The skull tends to smooth the scalp potential, compared to the brain surface or inner skull surface potential. Srinivasan et al. (1998) used spherical head models to quantify this effect. They generated random, many-dipole configurations in the cortex, and computed the scalp surface potentials. They sampled the scalp potential discretely using 19-, 32-, 64, and 128-channel arrays, and quantified the map differences for each array. They concluded that 128 electrodes are necessary to capture most of the spatial information available in the scalp potential, and that fewer than 64 channels can result in significant sampling errors. As in the time domain, if the scalp topography is sampled too sparsely, it suffers from aliasing artifacts. In the spatial domain, however, aliasing due to under-sampling can not be corrected by pre-filtering, as is done in the time domain.

7 Lead Field Theory

This section describes a useful way of thinking about the spatial resolution of scalp EEG. Previous sections described how each dipole (specified by position and orientation) gives a unique scalp potential. In this way of thinking, the potential for a single dipole is normally computed at all electrodes. Alternatively, the same problem may be arranged so that the potential difference across a single electrode pair is computed for each dipole position and orientations. This yields the *lead field vector* $\mathbf{L}$ for each electrode pair, which may be computed from the electric field that would exist in the head if current were injected into those same electrodes. This seems less intuitive but, insofar as scalp measurements integrate over the activity of large cortical areas (10–100 cm^2), this leads to a metric for the field of view of each electrode pair. The tabulation of the potential at every electrode, for each of a large but finite number of dipole locations and orientations in the brain, is called the *lead field matrix*. This quantity summarizes all the information about the head model, and is the starting point for inverse solutions.

7.1 Heuristic Definition

Imagine that a single dipole is placed at a point $\mathbf{r}_p$ inside a volume conductor, and oriented along the positive x-axis. Make no assumptions about the shape, homogeneity or isotropy of the volume conductor. Let $\Delta\Phi$ be the potential difference measured across two surface electrodes, and p_x be the dipole strength. Because Poisson's equation is linear in the sources, $\Delta\Phi$ must depend linearly upon the strength of the dipole, and this may be written algebraically as

$$\Delta\Phi = L_x p_x \tag{6.1}$$

where L_x is a proportionality constant. At the point $\mathbf{r}_p$, similar relations hold for the other two dipole orientations. If there were three perpendicular dipoles,

one along each of three Cartesian axes, then $\Delta\Phi$ would be the linear sum of each contribution.

$$\Delta\Phi = L_x p_x + L_y p_y + L_z p_z = \mathbf{L} \cdot \mathbf{p} \tag{6.2}$$

where the last equality derives simply from the definition of vector dot product. The quantity $\mathbf{L}$ is called the *lead field vector*. Strictly speaking we have not shown that $\mathbf{L}$ behaves as a vector under coordinate transformations, but it must if its contraction with the vector $\mathbf{p}$ is to yield a scalar $\Delta\Phi$.

7.2 Reciprocity Theorem

The reciprocity theorem (Helmholtz, 1853) gives an explicit expression for $\mathbf{L}$. The mathematical techniques used in deriving it are similar to those used in boundary element modeling. Consider a conducting volume V bounded by a surface S. Make no assumptions about the shape or homogeneity of the volume conductor.[5] Figure 4 shows two source and measurement configurations, denoted t_1 and t_2.

In configuration t_1, the source is a dipole located in the volume and the measurement is made by surface electrodes at positions $\mathbf{r}_\mathrm{A}$ and $\mathbf{r}_\mathrm{B}$. In configuration t_2, the source is introduced "reciprocally" by injecting current through the surface electrodes, and the potential difference is considered across the dipole. Now use Green's theorem to relate the potential Φ in one configuration to the current density $\mathbf{J}$ in the other. Consider the quantities

$$\vec{\nabla} \cdot [\Phi_1 \mathbf{J}_2] = \vec{\nabla}\Phi_1 \cdot \mathbf{J}_2 + \Phi_1 \vec{\nabla} \cdot \mathbf{J}_2$$
$$\vec{\nabla} \cdot [\Phi_2 \mathbf{J}_1] = \vec{\nabla}\Phi_2 \cdot \mathbf{J}_1 + \Phi_2 \vec{\nabla} \cdot \mathbf{J}_1$$

where $\mathbf{J}_i = -\sigma_i \vec{\nabla}\Phi_i$ for $i = 1, 2$. Subtracting these equations and assuming that $\sigma_1 = \sigma_2 = \sigma$, the first terms on the RHS of each equation cancel. By assumption $\vec{\nabla} \cdot \mathbf{J}_2 = 0$ in V and $\mathbf{J}_1 \cdot \hat{\mathbf{n}} = 0$ on S. Integrating over the volume V and using the divergence theorem to write the LHS as a surface integral over S leads to

$$\int_S \Phi_1 \, \mathbf{J}_2 \cdot \hat{\mathbf{n}} \, dS = - \int_V \Phi_2 \, \vec{\nabla} \cdot \mathbf{J}_1 \, dV \tag{6.3}$$

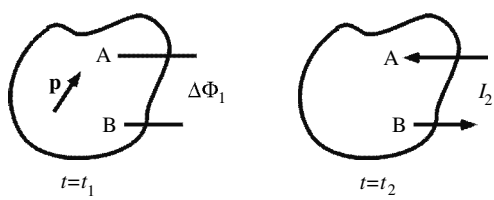


Fig. 4. Reciprocal source and measurement configurations for EEG reciprocity theorem

[5] The derivations presented here assume isotropy for simplicity.

In configuration t_1, let the current source and sink be located at $\mathbf{r}_\pm = \mathbf{r}_1 \mp \mathbf{d}/2$ and let the dipole strength be $p = I_1 d$, where d is the dipole separation. We have

$$\vec{\nabla} \cdot \mathbf{J}_1 = I_1 \left[\delta^{(3)}(\mathbf{r} - \mathbf{r}_+) - \delta^{(3)}(\mathbf{r} - \mathbf{r}_-)\right] \tag{6.4}$$

where the sign convention is such that $\mathbf{J}_1 = -\sigma_1 \vec{\nabla} \Phi_1$. In the notation of (3.12), $\vec{\nabla} \cdot \mathbf{J}_1 = \vec{\nabla} \cdot \mathbf{J}_\mathrm{E} = -\vec{\nabla} \cdot \mathbf{J}_\mathrm{S}$.

In configuration t_2, let $\mathbf{r}_\mathrm{A}$ be the location of the source electrode, which injects current into the head by establishing a positive potential at that point, and let $\mathbf{r}_\mathrm{B}$ be the location of the sink electrode, which extracts current from the head by establishing a negative potential at that point. The normal component of the current density on the surface may then be written formally

$$\mathbf{J}_2 \cdot \hat{\mathbf{n}} = I_2 \left[\delta^{(2)}(\mathbf{r} - \mathbf{r}_\mathrm{B}) - \delta^{(2)}(\mathbf{r} - \mathbf{r}_\mathrm{A})\right] \tag{6.5}$$

Inserting (6.4) and (6.5) into (6.3) and performing the integrals trivially over the delta functions gives

$$I_2 \left[\Phi_1(\mathbf{r}_\mathrm{A}) - \Phi_1(\mathbf{r}_\mathrm{B})\right] = -I_1 \left[\Phi_2(\mathbf{r}_+) - \Phi_2(\mathbf{r}_-)\right] \tag{6.6}$$

Expanding the difference $\Phi_2(\mathbf{r}_\pm)$ in powers of d and taking the usual dipole limit as $d \longrightarrow 0$ gives

$$\Phi_1(\mathbf{r}_\mathrm{A}) - \Phi_1(\mathbf{r}_\mathrm{B}) = \mathbf{p} \cdot \mathbf{L} \tag{6.7}$$

where the lead field vector is defined

$$\mathbf{L} = -\frac{\vec{\nabla} \Phi_2(\mathbf{r}_1)}{I_2} = \frac{1}{\sigma(\mathbf{r}_1)} \frac{\mathbf{J}_2(\mathbf{r}_1)}{I_2} \tag{6.8}$$

Thus the lead field vector $\mathbf{L}$ for a particular electrode pair (A,B) is proportional to the current density $\mathbf{J}_2$ which would be created in V at the dipole position $\mathbf{r}_1$ if unit current I_2 were injected through the electrode pair. The proportionality constant is the reciprocal of the local conductivity σ at the dipole location $\mathbf{r}_1$.

The lead field $\mathbf{L}$ has the content of the usual forward problem, but is interpreted somewhat differently. It is computed as a function of the dipole position for fixed electrode positions. That is opposite the normal formulation of the forward solution, in which the potential at any point is computed for fixed dipole location. In this way the lead field gives a measure of the sensitivity of a particular electrode pair to dipoles are arbitrary locations in the volume. This may be used to reduce the computational demand of the forward problem for a fixed electrode array.

7.3 Spatial Sensitivity of EEG

The lead field vector $\mathbf{L}$ is the proportionality constant between $\mathbf{p}$ and $\Delta\Phi$, and is a measure of the sensitivity of an electrode pair to dipoles at various

locations. Since the orientation dependence implemented by the dot product is rather trivial, the magnitude of the lead field vector $L \equiv |\mathbf{L}|$ may be defined as the sensitivity of an electrode pair (Rush and Driscoll 1968). The amount of tissue probed by a particular pair may be quantified through the concept of half-sensitivity volume (Malmivuo and Plonsey 1995; Malmivuo et al. 1997).

The half-sensitivity volume (HSV) is defined as follows. For a given electrode pair, we compute the scalar sensitivity $L(\mathbf{r})$ for many ($\sim 10^4$) points $\mathbf{r}$ inside the brain volume, and determine the maximum sensitivity $L_{\max}$ for this pair. We then identify all points in the brain volume whose sensitivity is at least $L_{\max}/2$. The HSV is the volume filled by these points. The threshold of $1/2$ is certainly arbitrary, but does give some indication of the volume in which the largest sensitivities occur. We further define the depth D of the sensitivity distribution as the maximum depth of all points included in the HSV. Using a four-sphere model of the human head, the outer radii of the four tissue layers are 8.0 cm (brain), 8.2 cm (CSF), 8.7 cm (skull) and 9.2 cm (scalp).

Figure 5 shows $\mathbf{L}$ in a two-dimensional plane including the electrodes (A,B) and the origin. The vector nature of $\mathbf{L}$ is retained to illustrate its dependence

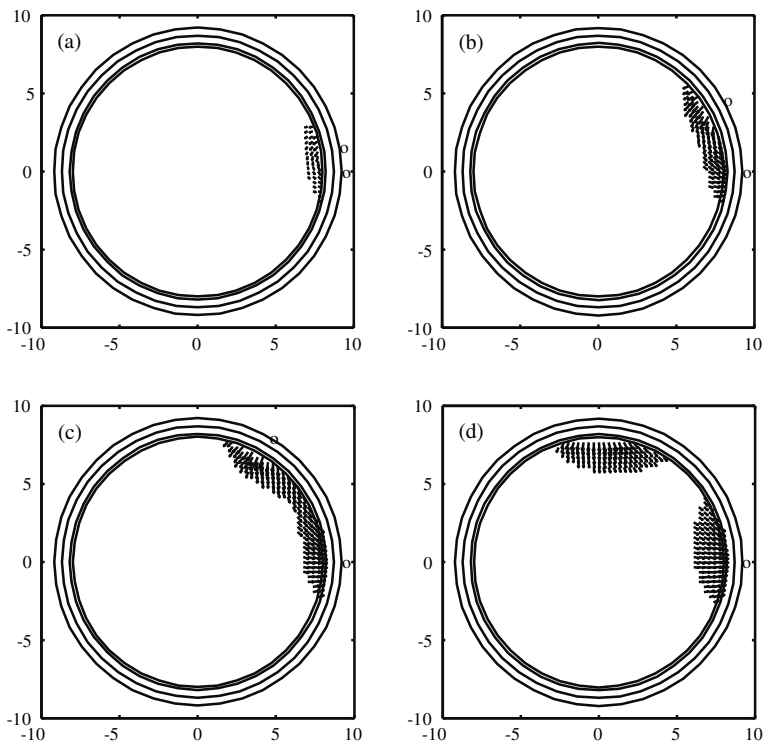


Fig. 5. The EEG lead field vector $\vec{L}(\vec{r})$ shown only within the HSV, for a four-sphere head model with $\sigma_3/\sigma_4 = 1/24$. The electrode separation angles θ are: (**a**) 10, (**b**) 30, (**c**) 60 and (**d**) 90 degrees. Axes are in cm

on orientation, but only its magnitude $L = |\mathbf{L}|$ is used to define the sensitivity and the HSV. In such a simple head model, the HSV is seen to be a single contiguous volume for nearby electrode pairs, which bifurcates near 60 degrees into two separate volumes for more distant pairs. Like the potential difference $\Phi_A - \Phi_B$, the lead field $\mathbf{L}$ changes only by a minus sign under interchange of A and B; the geometric pattern of sensitivity is unaffected.

The vector direction of $\mathbf{L}$ shows how the direction sensitivity of EEG bipolar recordings changes as a function of angle θ between the electrodes. Between nearby electrodes the sensitivity is primarily tangential to the sphere, while under each electrode the sensitivity is more radial. This observation refines the intuition that nearby electrodes are primarily sensitive to tangential dipoles between them. In fact, the greatest sensitivity lies not between the electrodes, but under each electrode, and has a significant radial component. For distant electrodes, the sensitivity is localized under each electrode separately. It is primarily radial, yet on the periphery of each lobe of the HSV there is some tangential component. This observation refines the intuition that distant electrodes are primarily sensitive to radial dipoles. In summary, both nearby and distant electrodes are sensitive to both radial and tangential dipoles. In both cases, the location of maximum sensitivity is directly under the electrodes, where the lead field $\mathbf{L}$ is oriented nearly radially. Thus EEG is predominantly but not exclusively sensitive to radial dipoles. This effect is enhanced by the fact that cortical gyri are populated with radial dipoles and are located closer to the detectors than are sulci.

Figures 6 and 7 show summarizations of the HSV results as a function of the angle θ between electrodes in the visualization plane. Intuitively, the smaller the HSV, the more refined an estimate of dipole position can be made

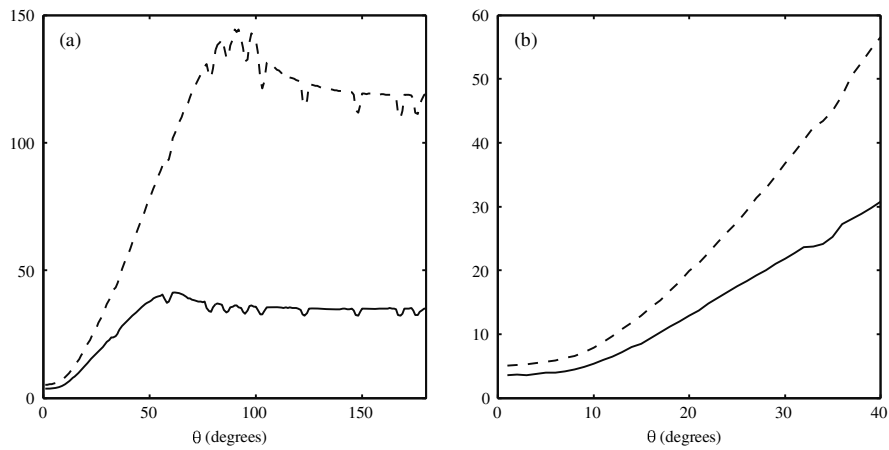


Fig. 6. Half-sensitivity volume (HSV) as a function of electrode separation angle θ, for $\sigma_3/\sigma_4 = 1/24$ (solid) and $\sigma_3/\sigma_4 = 1/80$ (dashed). Figure (**a**) is expanded in (**b**) for small θ

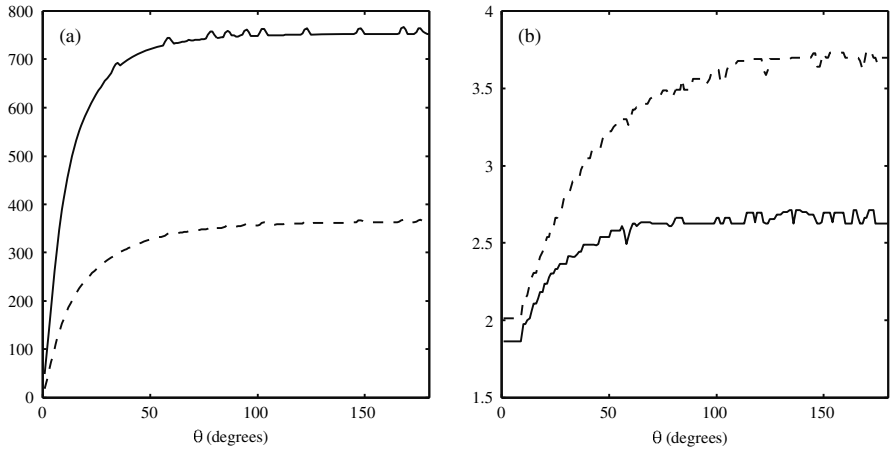


Fig. 7. (**a**) Maximum sensitivity $L_{\max}$ (Ω/m), and (**b**) depth of HSV (cm) as a function of electrode separation angle θ. Line types are the same as in Fig 2. Depth is defined relative to the scalp surface, which is separated from the brain surface by 1.2 cm

from a single electrode pair. Figure 6 shows the HSV as a function of angle. It increases rapidly as a function of angle until the bifurcation occurs, then decreases slightly. For very small θ, the HSV reaches an effective minimum. This limiting behavior can be understood in two complementary ways: In terms of the lead field vector computed via scalp current injection, for nearby electrodes most of the current is shunted through the scalp and little passes into the head volume. In terms of dipolar source currents, the brain potential is blurred by the skull such that nearby scalp electrodes sense nearly identical potentials. For conventional 19-electrode systems, for which nearby pairs are separated by 30 degrees, we find an optimal spatial resolution (minimum HSV) of 22–37 cm^3. For modern 129-electrode systems, for which nearby pairs are separated by more like 10 degrees, we find an optimal spatial resolution of 6–8 cm^3.

Two other sensitivity measures are maximum sensitivity and depth of sensitivity, shown in Fig.7. The maximum sensitivity (Fig. 7a) rises abruptly from near zero at small θ and approaches an asymptote. The maximum sensitivity is found at 180 degrees, and is 350–750 Ω/m depending on skull conductivity. The depth of sensitivity (Fig. 7b) varies similarly as a function of θ, with the exception of an abrupt minimum below 10 degrees. Like the small bumps visible in Figs. 6 and 7, the exact nature of this minimum appears to be artifactual, depending upon how the electrodes sit in relation to the Cartesian grid used to compute these quantities. The maximum depth is found at 180 degrees, and is 2.6–3.7 cm for this range of choices of skull conductivity.

Because of the folds of the cortical sheet, it is difficult to estimate the number of neurons detected by a particular electrode pair, without basing the analysis on a subject-specific structural MRI. In general, for nearby electrodes the HSV is confined almost entirely to the cortical sheet. Assuming the sheet is 0.25 cm thick and densely folded, these volume estimates above can be translated into effective cortical surface area estimates. Dividing the minimum HSV by the cortical thickness gives 88–148 cm^2 at 30 degrees, and 24–32 cm^2 at 10 degrees. Each 1 cm^2 of cortex is populated by approximately 10^7 pyramidal neurons (see Sect.2). This implies that on the order of 10^9 neurons reside inside the HSV at 30 degree electrode separation, and on the order of 10^8 at 10 degree electrode separation. These estimates are inflated, however, because some of the HSV includes noncortical tissue, and because cortical geometry excludes many neurons from detection when the local cortical surface is not parallel to the lead field. Nevertheless, these HSV measure provide a useful and intuitive metric of the spatial resolution of scalp EEG.

8 Topographic Analysis

As seen in Sects. 6 and 7, the reference electrode is an unavoidable fact in EEG recordings. Apart from attempts at optimal placement, several data processing methods exist for reducing or eliminating its influence. These include the average reference, the surface Laplacian, and inverse techniques which solve for brain dipole sources. The first two are specific to EEG, and avoid the ill-posed inverse problem. They make no explicit assumptions about the distribution of brain dipole sources, do not require head volume conductor models, and are computationally efficient. The average referenced scalp potential approximates the scalp potential referenced to infinity, and the surface Laplacian estimates the dura surface potential making the reasonable assumption of low skull conductivity. This section develops these ideas as simple and effective ways of handling the reference electrode issue in scalp EEG. Inverse methods based upon volume conductor models are discussed in the chapter by R. Leahy, and are applicable to both EEG and MEG.

8.1 EEG Reference Effects

At each time point, the definition of the electric potential by $\mathbf{E} = -\vec{\nabla}\Phi$ implies that Φ is ambiguous up to a constant. Physicists usually choose to reference Φ to infinity, so that the potential at infinity is zero by definition. This simplifies derivations and allows the potential at finite distances to be treated as a function of only one spatial variable. EEG recording systems with N amplifier channels record N potential differences from a common reference. If the reference electrode is located at the vertex, for example, then the potential differences measured at nearby electrodes will typically be smaller. Topographic maps of the raw potential, or derived quantities such as the Fourier power spectrum, will tend toward zero as the distance to the vertex is reduced,

The simplest attempt at eliminating the effect of the reference is to place it advantageously, i.e., away from active brain sources. Common choices include the earlobes, the nose, and the mastoids. The earlobes and nose are interesting, in light of the fact that electric potential tends to vary rapidly near pointed objects (Jackson 1975). The mastoids ensure secure attachment, which is most crucial for the reference electrode, but are clearly sensitive to brain activity in inferior posterior brain areas. Each of these are legitimate choices, although perhaps not effective in reaching their goal. Another approach, which should be avoided, is the linked-ears or linked-mastoids reference, in which electrodes are placed on both ears or mastoids, then physically linked before connecting to the reference input of the amplifier. EEG amplifiers are designed with high input impedances, specifically so they will not permit significant current flow across the scalp-electrode boundary. This reference choice violates that basic design principle, and leads to major problems. First, the linking introduces an highly conducting pathway between the two ears. This forces the ears to have similar potentials, which would not otherwise be the case for any singular choice of reference. In the limit of low scalp-electrode impedances, which is always the goal of electrode attachment, the potentials at the two reference sites are identical. Second, the impedances of the two reference electrodes are unlikely to be identical, so this choice is unlikely to be symmetrical as intended. Third, by violating the assumption of zero normal current flow though the scalp, the data are not suitable for analysis by the many commercial and open-source software packages. Fourth, because the basic physics of the problem has been altered, it is not possible simply to re-reference the data to other single electrodes.

8.2 Average Reference

The *average reference* is a simple way of estimating the potential at the reference electrode relative to infinity (Nunez 1981). At each time point, this quantity is used to compute the potentials at each measurement electrode relative to infinity (Bertrand et al. 1985). Because the genuine average reference can not be determined precisely, the operational average reference (based on limited sampling) has received valid criticism (Tomberg et al. 1990; Desmedt and Tomberg 1990) in favor of explicit references (Gencer et al. 1996; Geselowitz 1998), the surface Laplacian (Hjorth 1975; Nunez 1981), or more complicated methods (Lehmann et al. 1986; Yao 2001; Orekhova et al. 2002). Still it remains a useful technique for data analysis and visualization.

Biased Estimate

Let $\Phi(\mathbf{r})$ denote the scalp potential at point $\mathbf{r}$ measured relative to infinity, i.e., the *absolute* scalp potential. Let V_i denote the scalp potentials measured at electrodes $i = 1, ..., N$. The last electrode $i = N$ is the reference electrode for which $V_N \equiv 0$. For a perfect EEG amplifier system, we have $V_i = \Phi_i - \Phi_{\text{ref}}$,

where $\Phi_i = \Phi(\mathbf{r}_i)$, and $\Phi_N = \Phi_{\text{ref}}$ is the absolute potential at the reference electrode. We seek Φ_i but measure V_i; the difference amounts to estimating Φ_{ref}.

Let $\bar{V}$ denote the average of the potentials measured at N scalp electrodes:

$$\bar{V} \equiv \frac{1}{N}\sum_{i=1}^{N} V_i \tag{10.1}$$

Let U_i denote the average referenced potentials, i.e., re-referenced according to the definition $U_i \equiv V_i - \bar{V}$. The U_i have the property

$$\frac{1}{N}\sum_{i=1}^{N} U_i = \frac{1}{N}\sum_{i=1}^{N}\left(V_i - \bar{V}\right) = \left(\bar{V} - \bar{V}\right) = 0 \tag{10.2}$$

Because the sum over the U_i vanishes like the surface integral of Φ, the U_i are taken to estimate the Φ_i, with $\Phi_{\text{ref}} \equiv \Phi_N \simeq U_N = -\bar{V}$. This estimate is biased by not including contributions from the inferior head surface: the polar average reference effect (Junghofer et al. 1999).

Unbiased Estimate

Spherical splines were developed for topographic mapping of the scalp surface potential and the surface Laplacian (Perrin et al. 1989; Perrin et al. 1990), but their mathematical form carries implicitly an estimate of the average surface potential. Let $V(\mathbf{r})$ be the potential at an arbitrary point r on the surface of a sphere of radius r, and let $\mathbf{r}_i$ be the location of one the i^{th} measurement electrode. Spherical splines represent the potential at r on the surface of the sphere by

$$V(\mathbf{r}) = c_0 + \sum_{j=1}^{N} c_i \; g_m\left(\cos(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}_j)\right) \tag{10.3}$$

where the function $g_m(x)$ is given by

$$g_m(x) = \frac{1}{4\pi}\sum_{n=1}^{\infty} \frac{2n+1}{(n(n+1))^m} P_n(x) \tag{10.4}$$

The functions $P_n(x)$ are the Legendre polynomials of order n, which form a complete set of basis functions on a spherical surface.[6]

[6] The use of ordinary Legendre polynomials does not imply that the surface potential must have azimuthal symmetry. The variable x in $P_n(x)$ represents the angle between electrode position $\mathbf{r}_i$ and the interpolation point $\mathbf{r}$, so the claim is that (10.3) is capable of fitting the net scalp potential without inherent symmetry.

Recently we elaborated the idea put forth in Junghofer et al. (1999) that the spherical splines permit a better estimate of the average surface potential (Ferree 2006). Integrating (10.3) over the entire spherical scalp surface, and using that the integral of $P_n(x)$ on $-1 \leq x \leq +1$ vanishes for $n \neq 0$ (Arfken 1995), leads to

$$c_0 = \frac{1}{4\pi r_4^2} \int V(\mathbf{r})\ dS \tag{10.5}$$

where r_4 is the outer scalp radius. Thus the coefficient c_0 is equal to the average of the *interpolated* potential over the sphere surface.

Current conservation implies that, for dipolar current sources in an arbitrary volume conductor, the surface integral of the absolute potential Φ vanishes (Bertrand et al. 1985). Substituting $V(\mathbf{r}) = \Phi(\mathbf{r}) - \Phi_{\text{ref}}$ leads to

$$c_0 = \frac{1}{4\pi r_4^2} \int \left(\Phi(\mathbf{r}) - \Phi_{\text{ref}}\right)\ dS \simeq -\Phi_{\text{ref}} \tag{10.6}$$

Based upon (10.6), we expect c_0 to provide a *reasonable* estimate of Φ_{ref}, which can be used to compute the absolute potentials using $\Phi_i = V_i + \Phi_{\text{ref}} \simeq V_i - c_0$. This favorable situation is limited by the fact that the spline fit is severely under-constrained on the inferior head surface, and is unlikely to be numerically accurate there. It is conceivable that the estimate $\Phi_{\text{ref}} \simeq -c_0$ is *worse* than the usual estimate $\Phi_{\text{ref}} \simeq -\bar{V}$, but further investigation proved otherwise. A more convincing theoretical argument and numerical simulations showing that spherical splines generally provide a *better* estimate of Φ_{ref} are given in (Ferree 2006).

8.3 Surface Laplacian

Complementary to the scalp potential is the scalp surface Laplacian, the second spatial derivative of the potential. Practically speaking, the surface Laplacian solves the problem of the reference electrode because the second spatial derivative discards any overall constant (corresponding to the potential at the reference electrode relative to infinity). The calculation of the surface Laplacian is made separately at each time point. Physically, it is most directly related to the local current density flowing radially through the skull into the scalp. Because current flow through the skull is mostly radial, the scalp surface Laplacian remarkably provides an estimate of the dura potential (Nunez 1987). Numerical simulations using real data have shown that the surface Laplacian has 80–95% agreement with other dura imaging algorithms (Nunez and Srinivasan 2006). This connection between the scalp surface Laplacian and dura potential is derived next.

The following derivations make three main assumptions: 1) the current flow through the skull is nearly radial, 2) the potential drops across the scalp and CSF are small, at least compared to that across the skull, and 3) the

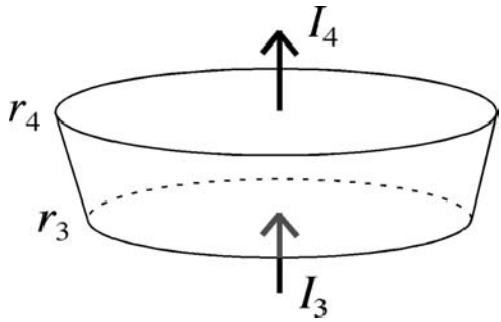


Fig. 8. A patch of scalp for consideration of the surface Laplacian in Problem 1. The parameter r_3 represents the outer skull surface, and r_4 the outer scalp surface. Alternatively, by replacing $3 \rightarrow 2$ and $4 \rightarrow 3$, the same figure may be used to represent a patch of skull in Problem 2

potential on the brain surface is much larger in amplitude than that on the scalp surface, by close proximity to the dipolar sources. Referring to Fig. 8, we have

$$I_3 = \int_S \mathbf{J}_4 \cdot \hat{t}\, d\Gamma \tag{10.7}$$

where Γ is the surface on the sides of the scalp patch, and $\hat{t}$ is a unit vector normal to Γ and therefore tangential to the scalp-air boundary. Assume that $\Phi_4(r, \theta, \phi)$ depends negligibly on r (on the grounds that the scalp is thin and σ_4 is high, at least compared to σ_3), so that $\Phi_4(r, \theta, \phi) \simeq V(\theta, \phi)$ leads to

$$I_3 \simeq -\sigma_4(r_4 - r_3)A_4\nabla_s^2 V \tag{10.8}$$

where A_4 is the cross-sectional area of the scalp patch. The boundary condition on $J_\perp$ on each side of the skull implies that the current flow through the skull is primarily radial, thus $I_2 = I_3$. Given that, the potential within the skull patch must vary radially according to the function

$$\Phi_3(r) = \frac{a}{r} + b \tag{10.9}$$

Considering how the cross-sectional area of the patch $A(r)$ varies as a function of r, and making use of the boundary condition on $J_\perp$ at the skull-scalp boundary r_3, shows that the potential difference across the skull is given approximately by

$$\Phi_3(r_2) - \Phi_3(r_3) = (r_3 - r_2)\frac{I_3}{\sigma_3}\frac{r_3}{r_2}\frac{1}{A_3} \tag{10.10}$$

Making use of the boundary condition on Φ leads to

$$\Phi_2(r_2) - \Phi_4(r_3) = -\frac{\sigma_4}{\sigma_3}\frac{r_4^2}{r_2 r_3}(r_4 - r_3)(r_3 - r_2)\nabla_s^2 V \tag{10.11}$$

which states that the potential difference across the skull is approximately proportional to the surface Laplacian. Finally, assuming that: 1) the potential drop across the CSF is small compared to that across the skull due to the low skull conductivity and high CSF conductivity, and the fact that the CSF layer is thinner than the skull, and 2) that the potential on the dura surface $\Phi_d = \Phi_2(r_1) = \Phi_1(r_1)$ is large compared to the potential on the scalp surface $\Phi_4(r_4) = V$, leads to

$$\Phi_d \simeq -\frac{\sigma_4}{\sigma_3}\frac{r_4^2}{r_2 r_3}(r_4 - r_3)(r_3 - r_2)\nabla_s^2 V \tag{10.12}$$

Thus the scalp surface Laplacian is proportional to the dura potential. Because the scalp surface Laplacian acts as a spatial high-pass filter (Nunez and Srinivasan 2006), possibly missing some valid information in the data, it is best used in conjunction with the average-referenced potential to study brain dynamics on the scalp.

8.4 Bipolar Pairs

Another way of eliminating the reference electrode effect is to based studies on bipolar pairs, as is common in clinical practice. Figs. 5(a) and (b) show that the potential difference between nearby electrode pairs have spatial sensitivity that is restricted to their local. The potential difference between all such pairs, or perhaps only nearest-neighbor pairs, may be computed easily and completely eliminates the dependence on the original reference electrode. Whereas the previous two approaches, the average reference and surface Laplacian, eliminated the dependence on the reference electrode, this approach makes explicit use of the reference electrode by effectively moving it around to form local bipolar pairs. Time-domain averages (i.e., event-related potentials) or power spectra computed from these time series are representative of the associated HSV, although the results are difficult to show graphically because each temporal or spectral measure is associated with one electrode rather than two. Time series collected from two bipolar pairs, which are themselves widely separated (e.g., a nearby pair in occipital cortex and a nearby pair in frontal cortex) may also be used for coherence analysis (Nunez 1995).

9 Summary

This goal of this chapter is to provide a rigorous introduction to scalp EEG for research in functional connectivity. We started at the microscopic level and discussed the cellular basis of current sources that generate extracellular fields, and developed the steps in electromagnetic theory that describe macroscopic fields in biological systems. We discussed the solutions to the EEG forward

problem in spherical and realistic head models. We also discussed EEG measurement technology, to make clear the reasons why the reference electrode issue arises so frequently in EEG experiments. We developed the concept of the lead field vector **L** to help visualize the spatial sensitivity patterns of scalp electrode measurements. These arguments lead to the conclusion that the reference electrode acts as a measurement electrode, and this fact must be addressed before drawing conclusions about the activity under any single electrode.

Studies of functional connectivity involve temporal measures of correlation, e.g., coherence and Granger causality, applied to two or more electrodes. Implicitly it is assumed that the time series collected at each electrode detects brain activity near that electrode. Our arguments using lead field theory show that each electrode is sensitive to large tissue volumes, containing perhaps 10^8–10^9 cortical neurons. Thus EEG measures of functional connectivity apply only to very large spatial scales, although somewhat smaller scale connectivity may be estimated with high resolution EEG methods like the surface Laplacian.

The reference electrode continues to confound many EEG studies. This chapter presented three practical ways of dealing with the reference electrode issue: adopting the average reference, the scalp surface Laplacian, or bipolar pairs. These data transformations and related concepts are essential to the estimation of temporal and spectral measures that may be used to make inferences about functional connectivity. Other facets of these topics are described elsewhere (e.g., Nunez and Srinivasan 2006).

Acknowledgements

This work was supported in part by NIH grants R43-MH-53768 and R43-NS-38788, and the Department of Radiology, University of California, San Francisco. The authors thank Matthew Clay for the numerical simulations in Section 7.3.

Index Words

EEG, MEG, membrane, cable theory, synchrony, electromagnetism, source current, return current, volume conduction, multipole expansion, dipole, head model, conductivity, boundary conditions, lead field, half-sensitivity volume, data recording, reference electrode, average reference, spline interpolation, surface Laplacian, spatial filter.

References

Arfken GB, Weber HJ (1995) *Mathematical Methods for Physicists.* Academic Press.

Barnard ACL, Duck IM, Lynn MS, Timlake WP (1967a) The application of electromagnetic theory to electrocardiography: I. Derivation of integral equations. Biophysical Journal 7: 443–462.

Barnard ACL, Duck IM, Lynn MS, Timlake WP (1967b) The application of electromagnetic theory to electrocardiography: II. Numerical solution of the integral equations. Biophysical Journal 7: 463–491.

Barr RC, Ramsey M, Spach MS (1977) Relating epicardial to body surface potential distributions by means of transfer coefficients based on geometry measurements. IEEE Trans. on Biomed. Eng. 24: 1–11.

Baumann SB, Wonzy DR, Kelly SK, Meno FM (1997) The electrical conductivity of human cerebrospinal fluid at body temperature. IEEE Trans. on Biomed. Eng. 44(3): 220–223.

Braitenberg V, Schuz A (1991) *Anatomy of the Cortex: Statistics and Geometry.* Springer-Verlag.

Elul E (1972) The genesis of the EEG. Int. Rev. Neurobiol. 15: 227–272.

Ferree TC, Luu P, Russell GS, Tucker DM (2001) Scalp electrode impedance, infection risk, and EEG data quality. Clinical Neurophysiology 112:536–544.

Ferree TC (2006) Spherical splines and average referencing in scalp electroencephalography. *Brain Topography* 19(1-2): 43–52.

Foster KR, Schwan HP (1989) Dielectric properties of tissues and biological materials: A critical review. Critical Reviews in Biomed. Eng. 17(1): 25–104.

Geddes, L. A. and L. E. Baker (1967). The specific resistance of biological materials: A compendium of data for the biomedical engineer and physiologist. Med. Biol. Eng. 5: 271–293.

Geselowitz DB (1967) On bioelectric potentials in an inhomogeneous volume conductor. Biophysical Journal 7: 1–11.

Geselowitz DB (1998) The zero of potential. *IEEE Eng. Med. Biol. Mag.* 17(1): 128–132.

Gulrajani RM (1998) *Bioelectricity and Biomagnetism.* John Wiley and Sons.

Helmholtz HLF (1853) Ueber einige Gesetze der Vertheilung elektrischer Strome in korperlichen Leitern mit Anwendung auf die thierisch-elektrischen Versuche. Ann. Physik und Chemie 89: 211–233, 354–377.

Huhta JC, Webster JG (1973) 60-Hz interference in electrocardiography. IEEE Transactions on Biomedical Engineering 20: 91–101.

Jackson JD (1975) *Classical Electrodynamics.* John Wiley and Sons.

Koch C, Segev I (1989) *Methods in Neuronal Modeling: From Synapses to Networks.* MIT Press.

Law SK, Nunez PL, Wijesinghe RS (1993) High-resolution EEG using spline generated surface Laplacians on spherical and ellipsoidal surfaces. IEEE Transactions on Biomedical Engineering 40(2): 145–153.

Law SK (1993) Thickness and resistivity variations over the upper surface of the human skull. *Brain Topography* 6(2): 99–109.

Malmivuo J, Plonsey R (1995) *Bioelectromagnetism.* Oxford University Press.

Malmivuo J, Suihko V, Eskola H (1997) Sensitivity distributions of EEG and MEG measurements. IEEE Transactions on Biomedical Engineering 44(3): 196–208.

Makeig S, Westerfield M, Jung TP, Enghoff S, Townsend J, Courchesne E, Sejnowski TJ (2002) Dynamic brain sources of visual evoked responses. Science 295(5555): 690–4.

Nunez PL (1981) *Electric Fields of the Brain*. Oxford University Press.

Nunez PL (1995) *Neocortical Dynamics and Human EEG Rhythms*. Oxford University Press.

Nunez PL, Srinivasan R (2005) *Electric Fields of the Brain*. 2nd Edition, Oxford University Press.

Oostendorp TF, Delbeke J, Stegeman DF (2000) The conductivity of the human skull: Results from *in vivo* and *in vitro* measurements. IEEE Trans. on Biomed. Eng. 47(11): 1487–1492.

Perrin F, Pernier J, Bertrand O, Echallier JF (1989) Spherical splines for scalp potential and current density mapping. Electroencephalography and Clinical Neurophysiology 72: 184–187.

Perrin F, Pernier J, Bertrand O, Echallier JF (1990) Corrigenda: EEG 02274, Electroencephalography and Clinical Neurophysiology 76: 565.

Plonsey R (1969) *Bioelectric Phenomena.* Mc-Graw-Hill.

Plonsey R (1982) The nature of sources of bioelectric and biomagnetic fields. Biophys. J. 39: 309–312.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) Numerical Recipes in C. Cambridge University Press.

Rush S, Driscoll DA (1968) Current distribution in the brain from surface electrodes. Anesthesia and analgesia 47(6): 717–723.

Rush S, Driscoll DA (1969) EEG electrode sensitivity – An application of reciprocity. IEEE Trans. on Biomed. Eng. 16(1): 15–22.

Schwan HP, Kay CF (1957) The conductivity of living tissues. Annals of New York Academy of Sciences 65: 1007.

Srinivasan R, Nunez PL, Tucker DM, Silberstein RB, Cadusch PJ (1996) Spatial sampling and filtering of EEG with spline Laplacians to estimate cortical potentials. Brain Topography 8(4): 355–366.

Srinivasan R, Tucker DM, Murias M (1998) Estimating the spatial Nyquist of the human EEG. Behavioral Research Methods, Instruments and Computers 30(1): 8–19.

Stok CJ (1987) The influence of model parameters on EEG-MEG single dipole source estimation. IEEE Trans. on Biomed Eng. 34(4): 289–296.

Zilles K (1990) Cortex. In: *The Human Nervous System* Pixinos G (ed.), Academic Press, New York.

Functional Imaging of Brain Activity and Connectivity with MEG

Felix Darvas and Richard M Leahy

Signal & Image Processing Institute, University of Southern California, Los Angeles CA90089

We present a survey of imaging and signal processing methods that use data from magnetoencephalographic (MEG) or electroencephalographic (EEG) measurements to produce spatiotemporal maps of neuronal activity as well as measures of functional connectivity between active brain regions. During the course of the chapter, we give a short introduction to the basic bioelectromagnetic inverse problem and present a number of methods that have been developed to solve this problem. We discuss methods to address the statistical relevance of inverse solutions, which is especially important if imaging methods are used to compute the inverse. For such solutions, permutation methods can be used to identify regions of interest, which can subsequently be used for the analysis of functional connectivity. The third section of the chapter reviews a collection of methods commonly used in EEG and MEG connectivity analysis, emphasizing their restrictions and advantages and their applicability to time series extracted from inverse solutions.

1 The Inverse Problem in MEG/EEG

Magnetoencephalography (MEG) measures non-invasively the magnetic fields produced by electrical activity in the human brain at a millisecond temporal resolution. The generators of these magnetic fields are dendritic currents in the pyramidal cells of the cerebral cortex. Since the currents produced by individual neurons are exceedingly weak, thousands of neurons have to be coherently active to produce a field that can be measured by MEG. The macroscopic fields generated by such ensembles of coherent neurons have strengths on the order of a few picotesla and are still one billion times smaller than the magnetic field of the earth.

A common electrical model for an ensemble of coherent neurons is the equivalent current dipole (ECD), which idealizes the ensemble as a single point source of electrical current. Due to the columnar organization of the cortex the ECD can be assumed to be oriented normally to the cortical surface

(Dale, and Serano, 1993, Okada, et al., 1997) and its location to be constrained to cortex. However, the model of an ECD with free orientation, which can be located anywhere within in the brain volume, remains common (Fuchs, et al., 1999).

The inverse problem is to find the neuronal activity, i.e. the location and strength of the associated ECDs, on the cerebral cortex or throughout the brain volume from noninvasive measurements of the magnetic fields produced outside the head. Likewise, if electroencephalographic (EEG) measurements are recorded, the change in scalp potentials due to the ECD inside the head volume is used to determine its strength and location. The solution of the inverse problem first requires the solution of a forward problem, which involves computation of the magnetic fields or electric potential changes outside the head due to an ECD in the brain volume. The basic physical laws from which the forward model for either MEG or EEG can be computed are Maxwell's equations under the assumption of stationarity (Hämäläinen, et al., 1993). This assumption is valid for the typical frequencies produced by the human brain, i.e. for frequencies on the order of 100 Hz, where the respective electromagnetic wavelengths (~ 300 m) far exceed the size of the head and thus changes in the fields produced by the neural currents inside the head can be considered instantaneous. Analytical solutions in geometries with spherical symmetry for the MEG/EEG forward problem have been discussed extensively by (Mosher, et al., 1999a, Zhang, 1995, Berg and Scherg, 1994, Sarvas, 1987). While for MEG the fields are only minimally distorted by biological tissue, anisotropies and inhomogeneities of the volume conductor have a strong impact on the scalp surface potentials. A simple spherical homogenous volume conductor model can be used for MEG with little impact on localization accuracy in the inverse solution (Leahy, et al., 1998), whereas for EEG it has been shown (Darvas, et al., 2006, Fuchs, et al., 2002, Baillet, et al., 2001), that numerical solutions of the forward model using a realistic head geometry can significantly improve the inverse solution over spherical models. Numerical methods such as the boundary element method (BEM), finite element method (FEM) and finite difference method (FDM) have been described in detail elsewhere (e.g. Fuchs, et al., 2001, Johnson, 1997). The forward model is solely dependent on the electromagnetic properties of the human head and the sensor geometry and is therefore independent of individual data recordings and need only be computed once per subject. Another important property of the forward model is that it is *linear* in terms of the strength of the neuronal currents. Consequently the summation of two source configurations produces the sum of the fields of the individual sources. For an individual ECD, the forward problem can be cast as a simple matrix-vector product as follows:

$$\underline{d}\,(t_i) = \underline{\underline{g}}\,(\underline{r}_k) \cdot \underline{q}\,(\underline{r}_k, t_i)\,, \underline{d} \in R^n, \underline{\underline{g}} \in R^{n\times 3}, \underline{q}, r \in R^3, \tag{1}$$

Where $\underline{d}$ is the vector of measurements collected for each of the n detectors, $\underline{\underline{g}}$ is the forward field for each detector for a source at location $\underline{r}_k$, and $\underline{q}$ is an ECD at location $\underline{r}_k$. For multiple sources eq. (1) can be expanded to

$\underline{d}(t_i) = \left[\underline{\underline{g}}(\underline{r}_1) \quad \underline{\underline{g}}(\underline{r}_2) \ \ldots \ \underline{\underline{g}}(\underline{r}_p)\right] \cdot \begin{bmatrix} \underline{q}(\underline{r}_1, t_i) \\ \underline{q}(\underline{r}_2, t_i) \\ \vdots \\ \underline{q}(\underline{r}_p, t_i) \end{bmatrix}$, and for multiple time points we can write the compact form:

$$\underline{\underline{D}} = \underline{\underline{G}} \cdot \underline{\underline{X}}, \underline{\underline{G}} \in R^{n \times 3p}, \underline{\underline{X}} \in R^{3p \times T}, \underline{\underline{D}} \in R^{n \times T}, \tag{2}$$

where p is the number of sources and T the number of time samples. The number n of detectors is typically on the order of 100, while the number of sources p can be either a few (1–10) when using dipole fitting methods or up to several thousand when using cortically constrained imaging methods.

Solutions to the inverse problem can be generally split into two classes: the dipole fitting or scanning solutions and imaging solutions (Darvas, et al., 2004). The aim of inverse methods is typically to minimize the cost function $\left\|\underline{\underline{D}} - \underline{\underline{G}} \cdot \underline{\underline{X}}\right\|_2^2$ with respect to the source parameters, subject to appropriate constraints on the sources $\underline{\underline{X}}$. In the case of dipole fitting, the solution is constrained to consist only of a limited number of sources, which leads to the least-squares dipole fit (Scherg, 1990). Optimizing the subspace correlation of the sources with the data and specifying a correlation threshold instead of the number of sources results in the RAP-MUSIC (recursively applied and projected multiple signal classification) solution of the inverse problem (Mosher, et al., 1999b). Distributed sources over the entire cortical surface or brain volume are allowed in the imaging approaches. To resolve ambiguities resulting from the larger number of sources, the least squares problem is regularized by addition of a norm on the image, which is selected to reflect the power, spatial smoothness or other characteristics of the solution, e.g. Hämäläinen, et al., 1993, Pascual-Marqui, et al., 1994. If the regularizer is quadratic then the solution is linear in the data and can be written:
$\underline{\underline{X}}^* = \left(\underline{\underline{G}}^t \cdot \underline{\underline{G}} + \underline{\underline{C}}^{-1}\right)^{-1} \underline{\underline{G}}^t \underline{\underline{D}}$, where the matrix $\underline{\underline{C}}$ encodes the source constraints. Assuming zero mean noise, the mean of the estimate $\underline{\underline{X}}^*$ can be related to the true source $\underline{\underline{X}}$ through the resolution kernel R: $E\left[\underline{\underline{X}}^*\right] = \underline{\underline{R}}\underline{\underline{X}}$, where $\underline{\underline{R}} = \left(\underline{\underline{G}}^t \cdot \underline{\underline{G}} + \underline{\underline{C}}^{-1}\right)^{-1} \underline{\underline{G}}^t \underline{\underline{G}}$. The mean of the estimate of the sources is therefore a linear combination of the true sources.

The beamformer methods (van Veen, et al., 1997, Robinson and Vrba, 1999) represent a hybrid approach to the inverse problem as they make assumptions of temporally independent dipolar sources, but they can also be used to compute activation images throughout the brain volume or on the cortical surface. Alternative solutions to the inverse problem have been proposed such as multipole fits (Jerbi, et al., 2004), which extend the elementary ECD source model to quadrupoles, a minimum norm type solution in a continuous source space (Riera, et al., 1998), and Bayesian and other nonquadratic regularization approaches that emphasize sparseness or other physiologically motivated properties in the solution as reviewed by Baillet, et al. (2001).

2 Statistical Significance of Inverse Solutions from Event Related Data

Once an inverse solution is computed, the question of the statistical significance of the solution arises. In most applications of MEG or EEG, single trial data are contaminated by environmental noise (power lines, electrical equipment), physiological noise (electrocardiogram, eye and muscle artifacts), and spontaneous brain activity. Consequently, meaningful source reconstruction usually requires some form of averaging of the measured *event related* brain activity over multiple repetitions (or epochs) of the same task. Typically, a pre-stimulus and a post-stimulus time segment is recorded, where the pre-stimulus segment is used to establish a baseline.

The inverse problem is solved either on the event related average of these single epochs or separately for each epoch, depending on the subsequent analysis to be performed. The single epochs, whose number can range from tens to hundreds, can also be used as the basis for assessing the significance of the inverse solution. If the solution is computed by a scanning or dipole fitting method, the significance of individual dipoles can be assessed in terms of the spatial accuracy with which the source is localized (Darvas, et al., 2005, Braun, et al., 1997). If the probability density function of the average data is known, bounds on the spatial confidence intervals for dipoles can be computed analytically by means of the Cramer-Rao lower bounds (Mosher, et al., 1993). Residuals can also be used to assess uncertainty with the chi-squared statistic. However, these methods are highly dependent on the assumed model and do not allow for factors such as missed sources or bias resulting from modeling errors. These problems can be avoided using a nonparametric resampling approach in which we use the single trial data to learn the distribution of the error. One example of such an approach is the Bootstrap method, which can be used to assess dipole localization uncertainty from single trial data (Darvas, et al., 2005). The Bootstrap method approximates the distribution of the data with the sample distribution formed by the collection of single epochs. A new representative sample of the data can then be generated by drawing a new set of single epochs at random and with replacement from the original epochs. Using this technique one can construct confidence intervals for each localized dipole, and reject those for which these intervals do not indicate a sufficiently accurate localization.

In the case of an imaging solution to the inverse problem, significance cannot be readily assessed in terms of localization uncertainty as sources can be reconstructed everywhere in the chosen source space (either cortical surface or brain volume). Instead, we can test at each voxel whether there is significant change in neuronal activity relative to a baseline condition. In more complex experimental designs, significant effects can similarly be assessed by fitting a general linear model at eachvoxel. This form of analysis is

very similar to that applied in detecting experimental effects in fMRI data. The problems differ in the degree of spatial correlation in the images and the intrinsic spatial resolution of the two modalities. However, in both cases we need to carefully control for the false positives that may result from performing multiple hypothesis tests (one per voxel). Thresholds for significant voxels in MEG linear minimum norm inverse solutions can be computed using either parameteric random fields methods or nonparametric permutation tests (Pantazis, et al., 2005). In both cases, voxelwise statistics are thresholded to control the family wise error rate (FWER), i.e. the probability of one or more false positives. Under the null hypothesis, the maximum distribution of a voxelwise statistic, computed over space or space and time, can be used to select a threshold for a desired FWER. For common distributions (Gaussian, chi-squared, student-t) and sufficiently smooth fields, the upper tail of the maximum distribution can be approximated using the expected value of the Euler characteristic (Worsley, et al., 1996). Application of this approach requires spatial smoothing of the brain activation maps to satisfy random field assumptions. Since resolution in MEG is already rather poor, it is preferable to avoid further smoothing. An alternative nonparameteric approach, which also avoids the need for smoothing, is to use permutations tests to learn the maximum distribution under the null hypothesis. The maximum distribution is learned by randomly permuting the post- and pre-stimulus segments of the epochs. Under the null hypothesis that there is no change in brain activation before and after the stimulus, these should be interchangeable and therefore suitable for generating a new sample of the average dataset. By applying the threshold to the reconstructed image, we create regions of interest that represent statistically significant activity (Fig. 1). Similar non-parametric permutation tests have also been applied to images obtained from beamformers (Singh, et al., 2003).

Time courses of activity for statistically selected regions of interest can be constructed by spatially integrating source activity over the region of interest or selecting the voxel that locally maximizes the test statistic.

In both dipole fitting and imaging solutions, one ends up with a number of localized regions in the brain, each showing significant event related activity. The time series from these locations can then be used for further processing to investigate functional connectivity.

If continuous data are recorded for which no event trigger is available, then the methods as described above cannot be applied. However, steady state activity under different test conditions can be compared using modifications of the thresholding methods described. Alternatively, if anatomical regions of interest are known a priori, then time series can be extracted from inverse solutions at these locations for processing using the methods described below.

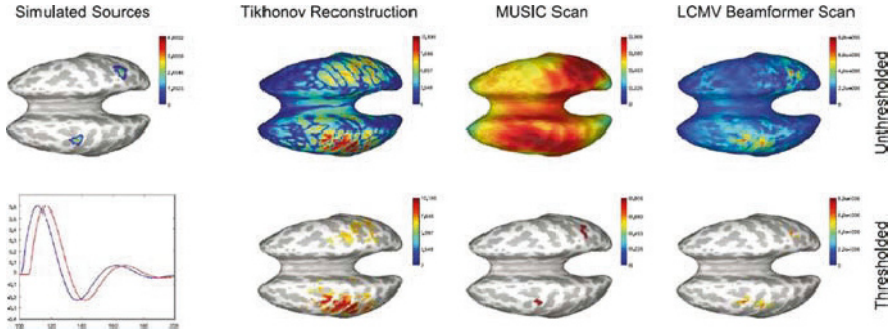


Fig. 1. Illustration of minimum norm, MUSIC and beamformer inverse methods for two simulated sources and the effect of thresholding on our ability to localize the sources (from Darvas, et al., 2004)

3 Assessing Functional Connectivity from the Inverse Solution

While MEG/EEG does not possess the good spatial resolution of fMRI, these modalities provide an excellent temporal resolution ($<$ 1ms), which fMRI cannot achieve as a result of the relatively slow hemodynamic response function (Kim, et al., 1997). This temporal resolution can be used to look at complex phenomena in the time-frequency domain, such as coherence, phase synchrony, or causality between signals from spatially separated brain regions. The goal of applying such measures is to establish functional correspondence between these regions in order to gain an understanding of the networks of neuronal populations that are involved in executing complex tasks. It should be noted that functional connectivity between regions merely establishes that these regions share mutual information during a specific task, whereas effective connectivity between regions actually implies a causal relationship between these regions (Friston 1994, Lee, et al., 2003, Horwitz, 2003).

Much of the work on functional connectivity using EEG or MEG (e.g. Gevins, et al., 1985, Simoes, et al., 2003) analyzes the connectivity between pairs of electrodes or channels directly. Due to the nature of the electromagnetic forward problem one has to deal carefully with *crosstalk* of sources. Because the fields or potentials created by an active neuronal source drop off at roughly $1/r^2$, where r is the distance from the source to the detector, neighboring detectors will typically be sensitive to the same sources, and in some cases all sensors may detect a single source. As described by eq. (2), the signals in each detector will be a weighted linear combination of the signals from all sources. It is possible to reduce sensitivity to crosstalk when analyzing sensor data, either by using connectivity measures that are insensitive to linear crosstalk or through application of statistical tests that allow for crosstalk under the null hypothesis. However, in principle it may be preferable to perform connectivity analysis in the source domain, since the

inverse solution can be seen as a spatial unmixing procedure. While the limited resolution of MEG/EEG means that inverse methods will not entirely remove crosstalk between sources, it will certainly be reduced. The maximum amount of crosstalk expected on average from linear inverse methods has been estimated to be less than 40% for EEG data in simulations (Liu, et al., 2002) and will be further reduced by combining MEG and EEG or by including prior information from fMRI. It should be noted that generally the crosstalk will be higher for sources which are located deeper in the brain due to the nonuniform resolution of the inverse solution, as reflected in the resolution kernel $\underline{\underline{R}}$ described above.

We now review a number of methods for assessing functional connectivity between brain regions, which are applicable to time series extracted from inverse MEG/EEG solutions.

Covariance

A straightforward approach to investigating dependence between regions is to look at the cross covariance or lagged cross covariance of two signals, the so called Evoked Potential Covariance (EPC) (Gevins and Cutillo, 1993, Smith 1992, Gevins and Bressler, 1988). The EPC is defined as the maximum covariance over all time lags. While this method can work well on regions that are sufficiently spatially separated, it will be seriously affected by crosstalk and will have very limited spatial resolution. This can be easily demonstrated in the following example:

Let $x_0(t)$ and $y_0(t)$ be two neuronal sources at locations r_x and $r_{y,}$, then the measurements in two channels $a(t)$ and $b(t)$ are given by $a(t) = g_a\left(\underline{r}_x\right) x_0\left(t\right) + g_a\left(\underline{r}_y\right) y_0\left(t\right)$, $b(t) = g_b\left(\underline{r}_x\right) x_0\left(t\right) + g_b\left(\underline{r}_y\right) y_0\left(t\right)$ and the channel covariance is given by

$$\begin{aligned}\operatorname{cov}(a,b) = {} & g_a\left(\underline{r}_x\right) g_b\left(\underline{r}_x\right) \operatorname{cov}(x_0,x_0) \\ & + g_a\left(\underline{r}_y\right) g_b\left(\underline{r}_y\right) \operatorname{cov}(y_0,y_0) \\ & + \left[g_a\left(\underline{r}_y\right) g_b\left(\underline{r}_x\right) + g_a\left(\underline{r}_x\right) g_b\left(\underline{r}_y\right)\right] \operatorname{cov}(x_0,y_0). \end{aligned} \tag{3}$$

Clearly, even if the sources themselves have zero covariance, the channel covariance will be non-zero, falsely implying a connectivity between the channels. It should be noted, however, that these false interactions can be corrected for in subsequent statistical analysis. Crosstalk can exist regardless of whether there is any true connectivity and therefore will also appear in data under the null hypothesis of no interaction. Consequently a suitable baseline condition can be used to control for false positives. Using a nonparameteric permutation test, it is straightforward to also correct for multiple comparisons (Blair and Karniski 1993).

Coherence

Coherence analysis of two signals is a similar approach in which the signals are represented in the frequency domain. Instead of looking for linear interactions between signals in the time domain, the cross spectrum of the two signals is computed and the coherence found as the magnitude of the cross spectrum normalized by the power spectra of the two signals. Event related coupling in the sensor domain has been widely reported (e.g. Mima, et al., 2001, Andres and Gerloff, 1999, Miltner, et al., 1999), but similarly to covariance analysis, is limited in spatial resolution due to the potential for crosstalk between detectors (Nunez, et al., 1997). Since the cross spectrum can be computed from the covariance of two signals by the Fourier transform, the cross spectrum between two channels can be found by the Fourier transform of eq. (3). Consequently, there is the possibility of falsely inferred interactions, unless controlled for in subsequent statistical analysis. Coherence analysis in the source domain has been used by Gross et al. in 2001 in their method for performing Dynamic Imaging of Coherent Sources (DICS) as the first stage of localizing coherent sources and extracting their time series for further analysis. In DICS the coherence of the sources is computed as the output of a frequency domain beamformer constrained to pass activity with unit gain at a specific location in a specific frequency band. The properties of this frequency domain beamformer are described in detail by Gross, et al. (2003).

The linearly constrained minimum variance (LCMV) beamformer used in DICS has the potential for partial cancellation of coherent sources, particularly if short segments of data are used to find the covariance matrix from which the beamformer weights are computed. Consequently, this method can give inaccurate results for strongly coherent sources. It should be noted, however, that the method is used by Gross et al, (2001) as a first step to localize candidate sources for a subsequent analysis of their *phase synchrony* (see section *phase locking*), and that in application of the method to experimental data the actual coherence can be expected to be relatively low ($<< 1$). Due to crosstalk effects, the original map will usually show a very high coherence value in the vicinity of the reference region and might be misleading. Consequently, when working with cortical coherence maps produced from DICS, it is useful to display the relative increase in coherence as compared to a baseline condition. An example of a typical DICS map with the baseline map subtracted is shown in Fig. 2. Coherence was computed in the 9–14 Hz band for MEG data recorded during a visually cued finger movement of the right index finger.

Since coherence mixes phase and power information, it might not be a suitable measure of connectivity between neuronal assemblies (David, et al., 2002, Le Van Quyen, et al., 2001, Rodriguez, et al., 1999). Furthermore, the definition of coherence through the cross spectral density requires second order stationarity, which is rarely the case for event related brain signals. We review alternative measures that emphasize phase coupling below.

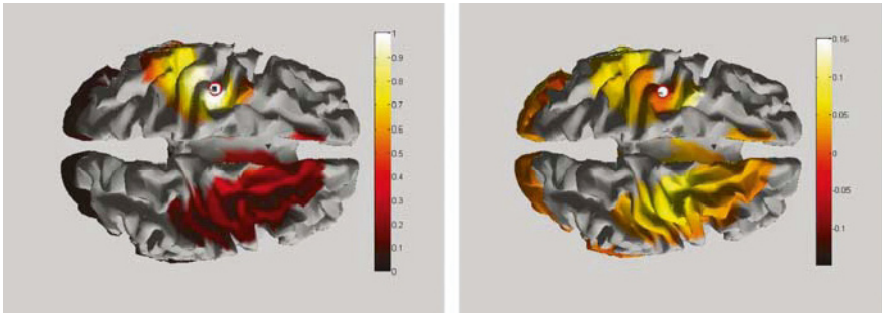


Fig. 2. Example of a DICS map from 275-channel MEG data for a motor activity experiment in the 9–14 Hz band, where the subject was performing a visually paced movement of the right index finger. The left image shows the coherence map relative to a source in the left M1 region. The right image shows a difference map in which the baseline coherence is subtracted. Both images were thresholded to control the false discovery rate of the DICS output at 1%

Coherency

Coherency is defined as the imaginary part of the cross spectrum and can be used to detect non-instantaneous interaction between two signals (Nolte et al., 2004). Let $x(t)$ and $y(t)$ be signals computed at two locations in the brain using one of the inverse methods discussed above and assume a set of sources $s_i(t)$ located within the brain. Because of limited resolution, $x(t)$ and $y(t)$ are a linear combination of the true signals. We denote the linear factors as a_i^x and a_i^y. The cross spectrum is then given by:

$$\begin{aligned} S_{xy}(f) &= \sum_{\tau} E\left[x(t)\, y(t+\tau)\right] e^{-j2\pi\tau\tau} \\ &= \sum_{\tau} \left[\sum_i \sum_j a_i^x a_j^y E\left[s_i(t)\, s_j(t+\tau)\right] \right] e^{-j2\pi f\tau} \end{aligned}$$

If the true signals $s_i(t)$ are uncorrelated, then the above reduces to $S_{xy}(f) = \sum_{\tau}\sum_i a_i^x a_i^y E\left[s_i(t)\, s_i(t+\tau)\right] e^{-j2\pi f\tau} = \sum_i a_i^x a_i^y S_{s_i s_i}(f)$, which is real valued as a result of the symmetry of the autocorrelation. A nonzero imaginary part in the cross spectrum must be due to something other than linear crosstalk and can be interpreted as a "true" interaction. Consequently, by considering the imaginary part only, or "coherency", one can eliminate the effect of crosstalk. While this approach can be used to detect interactions, quantitation is difficult since the real part of the cross spectrum, which contains information about both linear mixing and true interactions, is discarded.

Phase Locking

An alternative to the classical coherence measure is to separate phase and amplitude information in the signals and analyze the *phase locking* or *phase synchrony* as a measure of connectivity between neural assemblies. Phase locking between two oscillators is defined as $|\varphi_{n,m}(t)| < const, \varphi_{n,m}(t) = n\phi_1(t) - m\phi_2(t)$, where n and m are integers and $\phi_1(t)$ and $\phi_2(t)$ are the instantaneous phases of the signals (Tass, et al., 1998). In order to compute any measure of phase synchrony, the instantaneous phase of the signals must first be computed. This can be done by first narrowband filtering of the signals and then extracting the instantaneous phase by applying the Hilbert transform (Tass, et al., 1998). Alternatively phase can be computed from a time frequency decomposition using Gabor wavelets (Lachaux, et al., 1999). Once the instantaneous phase has been extracted from the signals, the *n:m synchronization index* (Tass, et al., 1998) or the *phase locking value* (PLV) (Lachaux, et al., 1999) can be computed from the phase difference of the signals. Tass et al. propose two methods to compute the synchronization index, one based on the Shannon entropy of the distribution of the phase differences and the other based on the conditional probability of the phases. The integer values n and m in this method are determined by trial and error, i.e. the synchronization index is computed for many combinations of n and m and those numbers which yield the largest index are selected for further analysis. Lachaux et al. define the PLV as $PLV_t = 1/N \left| \sum_n \exp\left(i\left(\phi_1\left(t,n\right) - \phi_2\left(t,n\right)\right)\right)\right|$, where the sum is taken over the number of trials. The PLV is also computed for surrogate data generated by shuffling the trial indices for one signal, while keeping the other constant, which allows for assessing the statistical significance of the observed synchronization. A comparison has shown that there is no fundamental difference between the synchronization index and PLV and that they are equally well suited for detecting phase synchrony (Le Van Quyen, et al., 2001). If applied in the channel domain, as noted by Lauchaux, et al., 1999, both methods also suffer from the crosstalk effect and can produce false synchrony measures, even in the absence of synchronous sources. Since sensors pick up signals from multiple sources and the phases of these sources do not combine linearly, crosstalk can result in a nonzero-phase difference, thus making it hard to distinguish from genuine phase locking.

Analysis of phase locking has revealed information about cortico-cortical and cortico-muscular coupling of signals (Fell, et al., 2001, Gross, et al., 2001, Gross, et al., 2000, Lauchaux, et al., 1999, Tass, et al., 1998). The limitation of phase locking methods is that they typically only consider interactions within a narrow frequency band. True large scale neural interactions may involve interactions between different frequencies, requiring consideration of a larger bandwidth (Lauchaux, et al., 1999, Tallon-Baudry, et al., 1997). An example showing the PLV between left and right motor cortices is shown in Fig. 3.

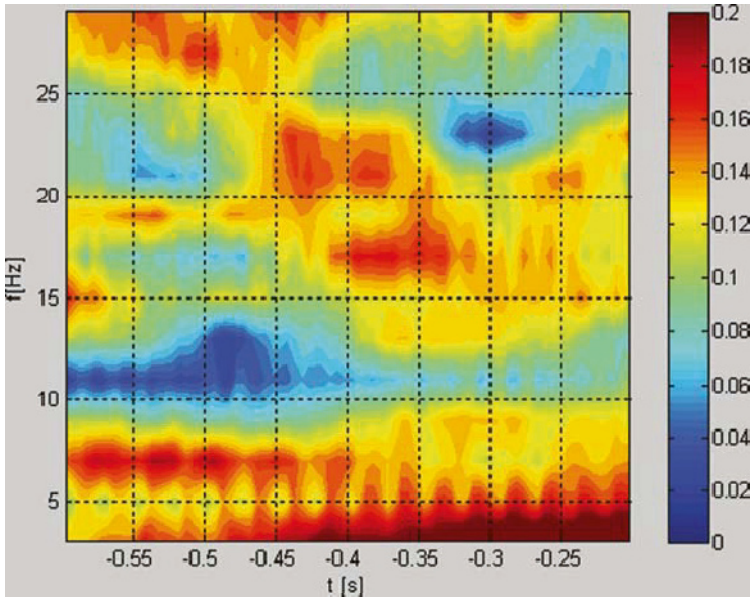


Fig. 3. Example of a PLV map computed using the wavelet transform for frequencies between 3 and 30 Hz. Phase locking was computed for the time series of the left and right motor cortices over a period of 400 ms prior to a movement of the right index finger. The time series were estimated from 275 channel MEG data using the minimum norm inverse

Higher Order Spectral Analysis

Higher order spectral analysis (Nikias and Mendel, 1993) can provide information about non-linear coupling between frequencies of different signals or within a single signal and can be seen as an extension of the classical power spectrum. While the power spectrum can be computed from the Fourier transform of the autocorrelation of the signal (and likewise, the cross spectrum from the Fourier transform of the cross correlation of two signals), higher order spectra are computed from the third, fourth or nth order cumulants of the signals. The *Bispectrum* for example is computed as the two-dimensional Fourier transform of the third-order cumulant of the signal. Let $x(k)$ be a n^{th} order stationary discrete time series, then the nth order cumulants are defined as

$$c_n(\tau_1, \tau_2, \ldots, \tau_{n-1}) = m_n(\tau_1, \tau_2, \ldots, \tau_{n-1}) - m_n^G(\tau_1, \tau_2, \ldots, \tau_{n-1})$$

where

$$m_n(\tau_1, \tau_2, \ldots, \tau_{n-1}) = E[x(k)\, x(k+\tau_1) \ldots x(k+\tau_{n-1})],$$

is the nth moment of x, and m_n^G is the respective moment of an equivalent Gaussian signal with the same mean and autocorrelation as x (Nikias and

Mendel, 1993). The nth order spectrum is thendefined as:

$$p_n(f_1, f_2, \ldots, f_{n-1}) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} \cdots \sum_{\tau_{n-1}=-\infty}^{\infty} c_n(\tau_1, \tau_2, \ldots, \tau_{n-1}) \times \exp\left(-i\left(\sum_{k=1}^{n-1} 2\pi f_k \tau_k\right)\right)$$

In a similar fashion, the nth order cross spectra can be computed by using the Fourier transforms of the cross-cumulants. For example, the bispectrum $p_3(f_1, f_2)$ between three signals is defined as the 2D Fourier transform of the cross-cumulants:

$$c_3^{xyz}(\tau_1, \tau_2) = E[x(k)\, y(k+\tau_1)\, z(k+\tau_2)] - m_3^{G^{xyz}}(\tau_1, \tau_2)$$

The cross-bicoherence can be computed by normalizing by the power spectral densities of the component signals: $C_d^{xyz}(f_1, f_2) = \frac{P_3^{xyz}(f_1,f_2)}{\sqrt{P_2^x(f_1)P_2^y(f_2)P_2^z(f_1+f_2)}}$. The bispectrum and cross-bispectrum, as well as the bicoherence and cross-bicoherence, can detect quadratic phase coupling between signals x, y and z. By replacing z with either x or y, we can use the cross bispectrum to detect coupling between any pair of signals as illustrated in Fig 4. Due to the asymmetry of the cross-bispectrum

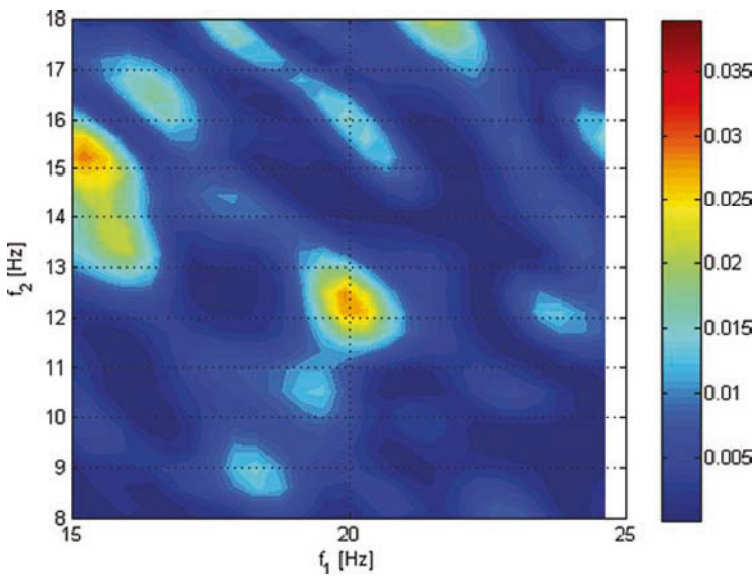


Fig. 4. Sample bicoherence map for the same time series described in Fig. 3. The map shows a non-linear interaction of 12 Hz and 20 Hz prior to the finger movement between the left and right motor cortex

for two signals, i.e. $C_2^{xxy}(f_1, f_2) \neq C_2^{yxx}(f_1, f_2)$, this measure can potentially be used to infer the direction of interaction from one signal to another.

Higher order spectra have many desirable properties that make them useful tools for analyzing neuronal interaction. The method is not limited to narrowband signals and the higher order spectra reveal phase information, which makes the method well suited to the analysis of time phase coupling (Jamsek, et al., 2003, Schack, et al., 2002). Because the higher order moments for a signal with a Gaussian distribution vanish, the higher order spectra are insensitive to Gaussian noise. Furthermore, in contrast to the coherence metric, the bispectrum is less sensitive to linear coupling and therefore is an attractive alternative for the detection of nonlinear interactions (Jamsek, et al., 2003). Applications of higher order spectral analysis to EEG data in the channel domain have been presented by Pfurtscheller and Lopes de Silva (1999), where non-linear interaction between 11 Hz and 22 Hz components in the post-movement beta event-related synchronization (ERS) was demonstrated, as well as in the analysis of short-term memory (Schack, et al., 2002) and in microelectrode recordings from the visual cortices of cats and monkeys (Schanze and Eckhorn, 1997). Because of its insensitivity to linear relationships and therefore its robustness to linear cross talk effects between signals, the method is also well suited for application in the source domain.

Structural Equation Modeling

While the methods described so far in this chapter can be used to analyze the interactions between two signals, it is also of great interest to perform network analysis involving multiple sources or ensembles of neuronal activity. A simple approximation to the potentially complex interactions between multiple regions is given by the *structural equation model* (SEM) or *path analysis*, which assumes a linear relationship between the activity in each region and also with respect to any experimental variables (Astolfi, et al., 2005, Friston, 1994,, Bollen 1989). This assumption can be cast in a simple equation (McIntosh and Gonzalez-Lima, 1994):

$$\begin{aligned} &\underline{y} = \underline{\underline{B}} \cdot \underline{y} + \underline{\underline{\Gamma}} \cdot \underline{x} + \underline{\zeta} \\ &\underline{y}, \underline{\zeta} \in R^m, \underline{x} \in R^n, \underline{\underline{B}} \in R^{n \times n}, \underline{\underline{\Gamma}} \in R^{m \times n} \end{aligned} \tag{4}$$

The vector $\underline{y}$ represents the signal at a single time point for each of the m connected components of the network and the vector $\underline{x}$ represents n independent external components, e.g. variations of the experimental conditions. These can be binary variables that are set to one for the experimental condition and zero for the control condition (McIntosh and Gonzalez-Lima, 1994). The matrices $\underline{\underline{B}}$ and $\underline{\underline{\Gamma}}$ represent, respectively, the influence of the network on itself and the influence of experimental conditions on the network. In SEM it

is assumed that the components of the network do not interact with themselves, therefore the diagonal elements of the matrix $\underline{\underline{B}}$ are set to zero. The vector $\underline{\zeta}$ models the residual activity not explained by the linear model. The model parameters are estimated from the observed signals using least squares regression (Astolfi et al., 2005). Since typically the number of unknowns in this optimization problem exceeds the number of equations, a priori information about the model has to be provided. This is done by limiting the number of *free parameters*, i.e. the elements of the matrix $\underline{\underline{B}}$. By setting some elements of $\underline{\underline{B}}$ to zero, interactions between selected components are precluded and the number of unknowns is reduced, making the system solvable. Because the method is based on the covariance it will be affected by linear crosstalk leading to potentially erroneous inferences of network connectivity when applied to EEG or MEG data. The SEM approach was originally applied in functional neuroimaging to PET and fMRI studies in which dynamic data were not available. However, when applied to dynamic EEG and MEG data, the assumption of instantaneous linear unidirectional interactions between the network components is rather restrictive. The use of only the zero-lag covariance between regions ignores the temporal structure of the data, so that shuffling the data in time would have no influence on the SEM parameters (Ramnani, et al., 2004). We conclude this chapter with a brief introduction to two network models that do consider dynamic interactions: the linear MVAR and the nonlinear dynamic causal models.

Multivariate Autoregressive Models (MVAR)

Multivariate autoregressive models can also be used to model interactions between multiple regions. However, unlike SEM, they do not assume instantaneous linear interaction between the regions, but also take the past of each signal into account. One of the key attractions of the MVAR model is its ability to make inferences about the direction of interaction of multiple network components (Brovelli, et al., 2004). The mth order MVAR model is described by the following equation:

$$\underline{x}(t) = \sum_{i=1}^{m} \underline{\underline{A}}_i \underline{x}(t-i) + \underline{e}(t)$$

The vector $\underline{x}(t)$ contains the time series for each region, the vector $\underline{e}(t)$ is a zero-mean uncorrelated noise process and the coefficient matrices $\underline{\underline{A}}_i$ can be found by solving the multivariate Yule-Walker equations. If transformed to the frequency domain, the MVAR equation takes on the more simple form $\underline{x}(f) = \underline{\underline{H}}(f) \cdot \underline{E}(f)$, where $\underline{\underline{H}}(f)$ is the transfer function matrix (Kus, et al., 2004). From the frequency domain representation we can compute the Directed Transfer Function (DTF), a frequency dependent measure of the interaction between nodes in the network (Kaminski, et al., 2001). Similarly to phase-locking and coherence measures, the DTF is a measure of

interaction at a single frequency, although these interactions are computed simultaneously across the entire frequency range of interest. An advantage of MVAR models is that they allow one to determine the direction of interaction by analogy to the concept of Granger Causality. Furthermore, unlike the SEM models, no a priori knowledge is required about connectivity in the network. The wide sense stationarity requirement does limit the utility of MVAR models in event related studies, but since they are parametric they can be fit using relatively few temporal samples (with the covariance computed by averaging across epochs) and nonstationarity then detected using a sliding window. Since the method is linear, crosstalk will have a significant effect on DTF. If the mixed signals are given by $\underline{y} = \underline{\underline{M}} \cdot \underline{x}$, then instead of the 'true' coefficient matrices $\underline{\underline{A}}_i$ one gets the mixed coefficient matrices $\underline{\underline{A}}_i\underline{\underline{M}}$, but there is no way to identify the mixing matrix $\underline{\underline{M}}$ from the estimated coefficient matrices.

Neural Mass Models

Neural mass models (David et al., 2004, David and Friston, 2003, Jansen and Rit, 1995) are physiologically motivated models of cortical activity, which attempt a mathematical description of the network of neurons that generate the signals measured by MEG/EEG. The purpose of these models is to provide a realistic network simulation of cortical activity with controllable connectivity parameters, but they can also be used to estimate network parameters from real data (David and Friston, 2003). The basic elements of the model are excitatory and inhibitory columns of neurons (Jansen and Rit, 1995). The state of each element is described by an average membrane potential and a mean firing rate. The membrane potential and average firing rate are related by nonlinear input/output functions for each element and the parameters of these functions are thus the parameters of the element. A number of interaction constants between the ensembles can be used to describe the network structure. If no external input signal is used, the model is typically driven by random noise. David and Friston (2003) expanded the double column model proposed by Jansen and Rit (1995) to a meta-ensemble or *mass model*, which instead of containing only two columns is comprised of N columns. These neuronal mass models can be thought of as cortical areas and by introduction of multiple areas and coupling between those, macroscopic behavior of the brain, as can be measured by MEG/EEG, can be simulated. Mass models can exhibit a broad range of frequencies and their non-linear nature and controllable macroscopic coupling make them good candidates for testing connectivity measures (David et al., 2004). Also by defining meta parameters of the model such as the number of areas and their connectivity, the models themselves can be used to fit real measurements and thus provide an estimate of functional connectivity (David and Friston, 2003).

References

Andres F, Gerloff C (1999) Coherence of Sequential Movements and Motor Learning. Journal of Clinical Neurophysiology 16(6), 520–527

Astolfi L, Cincotti F, Babiloni C, Carducci F, Basilisco A, Rossini PM, Salinari S, Mattia D, Cerutti S, Ben Dayan D, Ding L, Ni Y, He B, Babiloni F (2005) Estimation of the Cortical Connectivity by High-Resolution EEG and Structural Equation Modeling: Simulations and Application to Finger Tapping Data. IEEE Transactions on Biomedical Engineering 52(5), 757–768

Baillet S, Mosher JC, Leahy RM (2001) Electromagnetic Brain Mapping. IEEE Signal Processing Magazine 18 (6), 14–30

Baillet S, Riera JJ, Marin G, Mangin JF, Aubert J, Garnero L (2001) Evaluation of inverse methods and head models for EEG source localization using a human skull phantom. Physics in Medicine and Biology 46, 77–96

Berg P, Scherg M (1994) A fast method for forward computation of multiple-shell spherical head models. Electronencephalography and Clinical Neurophysiology 90, 58–64

Blair RC, Karninski W (1993) An alternative method for significance testing of waveform difference potentials. Psychophysiology 30(5), 518–524

Bollen KA (1989), Structural Equations with Latent Variables. Wiley. New York,

Braun C, Kaiser S, Kincses WE, Elbert T (1997) Confidence Interval of Single Dipole Locations Based on EEG Data. Human Brain Mapping 10(1), 31–39

Brovelli A, Ding M, Ledberg A, Chen Y, Nakamura R, Bressler SL (2004) Beta Oscillations in large-scale sensorimotor cortical network: Directional influences revealed by Granger causality. Proceedings of the National Academy of Sciences 101(26), 9849–9854

Dale AM, Serano MI (1993) Improved localization of cortical activity by combing EEG and MEG with MRI cortical surface reconstruction: a linear approach. Journal of Cognitive Neuroscience 5, 162–176

David O, Friston KJ (2003) A neural mass model for MEG/EEG:coupling and neuronal dynamics. Neuroimage 20, 1743–1755

David O, Cosmelli D, Friston KJ (2004) Evaluation of different measures of functional connectivity using a neural mass model. Neuroimage 21, 659–673

Darvas F, Pantazis D, Kucukaltun-Yildirim E, Leahy RM (2004) Mapping human brain function with MEG and EEG:methods and validation. Neuroimage 23S, 289–299

Darvas F, Rautiainen M, Pantazis D, Baillet S, Benali H, Mosher JC, Garnero L, Leahy RM (2005) Investigations of dipole localization accuracy in MEG using the bootstrap. Neuroimage 25, 355–368

Darvas F, Ermer JJ, Mosher JC, Leahy RM (2006) Generic Head Models for Atlas-Based EEG Source Analysis. Human Brain Mapping 27, 129–143

David O, Garnero L, Cosmelli D, Varela FJ (2002) Estimation of Neural Dynamics from MEG/EEG Cortical Current Density Maps: Application to the Reconstruction of Large-Scale Cortical Synchrony. IEEE Transactions on Biomedical Engineering 49(9), 975–987

Fell J, Klaver P, Lehnertz K, Grunwald T, Schaller C, Elger CE, Fernandez G (2001) Human memory formation is accompanied by rhinal-hippocampal coupling and decoupling. Nature neuroscience 4(12), 1259–1264

Friston KJ (1994) Functional and effective connectivity in neuroimaging: a synthesis. Human Brain Mapping 2, 256–278

Fuchs M, Wagner M, Köhler T, Wischmann HA (1999) Linear and Nonlinear Current Density Reconstructions. Journal of Clinical Neurophysiology 163(3), 267–295

Fuchs M, Wagner M, Kastner J (2001) Boundary element method volume conductor models for EEG source reconstruction. Journal of Clinical Neurophysiology 112(8), 1400–1407

Fuchs M, Kastner J, Wagner M, Hawes S, Ebersole JS (2002) A standardized boundary element method volume conductor model. Journal of Clinical Neurophysiology 113(5), 702–712

Gevins AS, Doyle JC, Cutillo BA, Schaffer RE, Tannehill RS, Bressler SL (1985). Neurocognitive pattern analysis of a visuospatial task: rapidly-shifting foci of evoked correlations between electrodes. Psychophysiology 22, 32–43.

Gevins AS, Bressler SL (1988) Functional topography of the human brain. In Pfurtscheller G (Ed.) Functional Brain Imaging. Hans Huber, Bern, 99–116

Gevins AS, Cutillo BA (1993) Spatiotemporal dynamics of component processes in human working memory. Electroencephalography and clinical Neurophysiology 87, 128–143

Gross J, Tass PA, Salenius S, Hari R, Freund HJ, Schnitzler A (2000) Corticomuscular synchronization during isometric muscle contraction in humans as revealed by Magnetoencephalography. Journal of Physiology 527(3), 623–631

Gross J, Kujala J, Hämäläinen M, Timmermann L, Schnitzler A, Salmelin R (2001) Dynamic imaging of coherent sources: Studying neural interactions in the human brain. Proceedings of the National Academy of Sciences 98(2), 694–699

Gross J, Timmermann L, Kujala J, Salmelin R, Schnitzler A (2003) Properties of MEG tomographic maps obtained with spatial filtering. Neuroimage 19, 1329–1336

Hämäläinen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV (1993) Magnetoencephalography – theory, instrumentation, and applications to noninvasive studies of the working human brain. Reviews of Modern Physics 65(2), 413–497

Horwitz B (2003) The elusive concept of brain connectivity. Neuroimage 19, 466–470

Jamsek J, Stefanovska A, McClintock PVE, Khovanov IA (2003) Time-phase bispectral analysis. Physical Review E 68, 016201 1–12

Jansen BH, Rit VG (1995) Electronencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biological Cybernetics 73, 357–366

Jerbi K, Baillet S, Mosher JC, Nolter G, Garnero L, Leahy RM (2004) Localization of Realistic Cortical Activity in MEG using Current Multipoles. Neuroimage 22(2), 779–793

Johnson CR (1997) Computational and numerical methods for bioelectric field problems. Critical Reviews in Biomedical Engineering 25(1), 1–81

Kaminski M, Ding M, Truccolo WA, Bressler SL (2001) Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. Biological Cybernectics 85(2), 145–157

Kim SG, Richter W, Ugurbil K (1997) Limitations of temporal resolution in functional MRI. Magnetic resonance in medicine 37(4), 631–636

Okada Y, Wu J, Kyuhou S (1997) Genesis of MEG signals in a mammalian CNS structure. Electroencephalography & Clinical Neurophysiology 103, 474–485

Kus R, Kaminski M, Blinowska KJ (2004) Determination of EEG activity propagation: pair-wise versus multi-channel estimate. IEEE Transactions on Biomedical Engineering 51(9), 1501–1510

Lachaux JP, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring Phase Synchrony in Brain Signals. Human Brain Mapping 8, 194–208

Leahy RM, Mosher JC, Spencer ME, Huang MX, Lewine JD (1998) A study of dipole localization accuracy for MEG and EEG using a human skull phantom. Electronencephalography and Clinical Neurophysiology 107(2), 159–173

Lee L, Harrison LM, Mechelli A (2003) A report on functional connectivity workshop, Düsseldorf 2002. Neuroimage 19, 457–465

Le Van Quyen M, Foucher J, Lauchaux JP, Rodriguez E, Lutz A, Martinerie J, Varela FJ (2001) Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony. Journal of Neuroscience Methods 111, 83–89

Liu AK, Dale AM, Belliveau JW (2002) Monte Carlo Simulation Studies of EEG and MEG Localization Accuracy. Human Brain Mapping 16, 47–62

McIntosh AR, Gonzalez-Lima F (1994) Structural Equation Modeling and Its Application to Network Analysis in Functional Brain Imaging. Human Brain Mapping 2, 2–22

Miltner W, Braun C, Arnold M, Witte H, Taub E (1999) Coherence of gamma-band EEG activity as a basis for associative learning. Nature 397,434–436

Mima T, Oluwatimilehin T, Hiraoka T, Hallet M (2001) Transient Interhemispheric Neuronal Synchrony Correlates with Object Recognition. Journal of Neuroscience 21(11), 3942–3948

Mosher JC, Spencer ME, Leahy RM, Lewis P (1993) Error bounds for EEG and MEG dipole source localization. Electroencephalography and Clinical Neurophysiology 86(5), 303–321

Mosher JC, Leahy RM, Lewis PS (1999a) EEG and MEG: forward Solutions for Inverse Methods. IEEE Transactions on Biomedical Engineering 46(3), 245–259

Mosher JC, Leahy RM (1999b) Source localization using recursively applied and projected (RAP) music. IEEE Transactions on Signal Processing 47(2), 332–340

Nikias CL, Mendel JM (1993) Signal Processing with Higher-Order Spectra. IEEE Signal Processing Magazine 10(3), 10–37

Nolte G, Bai O, Wheaton L, Mari Z, Vorbach S, Hallet M (2004) Identifying true brain interaction from EEG data using the imaginary part of coherency. Clinical Neurophysiology 115, 2292–2307

Nunez PL, Srinivasan R, Westdorp AF, Wijesinghe RS, Tucker DM, Silberstein RB, Cadusch PJ (1997) EEG coherency I:statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. Electronencephalography and clinical Neurophysiology 103, 499–515

Pantazis D, Nichols TE, Baillet S, Leahy RM (2005) A comparison of random field theory and permutation methods for the statistical analysis of MEG data, Neuroimage 25, 383–394

Pascual-Marqui RD, Michel CM, Lehmann D (1994) Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. International Journal of Psychophysiology 18(1), 49–65

Pfurtscheller G, Lopes da Silva FH (1999) Event-related EEG/MEG synchronization and desynchronization: basic principles. Journal of Clinical Neurophysiology 110, 1842–1857

Ramnani N, Behrens TEJ, Penney W, Matthews PM (2004) New Approaches for Exploring Anatomical and Functional Connectivity in the human Brain. Biological Psychiatry 56, 613–619

Riera JJ, Fuentes ME, Valdes PA, Oharriz Y (1998) EEG-distributed inverse solutions for a spherical head model. Inverse Problems 14, 1009–1019

Robinson SE, Vrba J.(1999). Functional neuroimaging by synthetic aperture magnetometry (SAM). Recent Advances in Biomagnetism. Tohoku Univ. Press, Sendai, 302–305.

Rodriguez E, George N, Lauchaux JP, Matinerie J, Renault B, Varela FJ (1999) Perception's shadow: long-distance synchronization of human brain activity. Nature 397, 430–433

Sarvas J (1987) Basic mathematical and electromagnetic concepts of the biomagnetic inverse problems. Physics in Medicine and Biology 32, 11–22

Schack B, Vath N, Petsche H, Geissler HG, Möller E (2002) Phase-coupling of theta-gamma EEG rythms during short-term memory processing. International Journal of Psychophysiology 44, 143–163

Schanze T, Eckhorn R (1997) Phase correlation among rythms present at different frequencies: spectral methods, application to microelectrode recordings from visual cortex and functional implications. International Journal of Psychophysiology 26, 171–189

Scherg M (1990) Fundamentals of dipole source potential analysis. In Grandori F, Hoke M, Romani GL (Ed.) Advances of Audiology. Karger, Basel, 40–69

Simoes C, Jensen O, Parkkonen L, Hari R (2003) Phase locking between human primary and secondary somatosensory cortices. Proceedings of the National Academy of Sciences 100(5), 2691–2694

Singh KD, Barnes GR, Hillebrand A (2003) Group imaging of task-related changes in cortical synchronization using nonparametric permutation testing. Neuroimage 19, 1589–1601

Smith WE (1992) Estimation of the Spatio-Temporal Correlations of Biological Electrical Sources from Magnetic Their Fields. IEEE Transactions on biomedical engineering 39(10), 997–1004

Tallon-Baudry C, Bertrand O, Delpuech C, Pernier J (1997) Oscillatory γ-Band (30–70 Hz) Activity Induced by a Visual Search Task in Humans. The Journal of Neuroscience 17(2), 722–734

Tass P, Rosenblum MG, Weule J, Kurths J, Pikovsky A, Volkmann J, Schnitzler A, Freund HJ (1998) Detection of n:m Phase Locking from Noisy Data: Application to Magnetoencephalography. Physical Review Letters 81(15), 3291–3294

Worsley KJ, Marret S, Neelin P, Vandal AC, Friston KJ, Evans AC (1996) A unified statistical approach for determining significant signals in images of cerebral activation. Human Brain Mapping 4, 58–73

Van Veen BD, van Drongelen W, Yuchtman M, Suzuki A (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Transactions on Biomedical Engineering 44 (9), 867–880

Zhang Z (1995) A fast method to compute surface potentials generated by dipoles within multilayer anisotropic spheres. Physics in Medicine and Biology 40, 335–349

Insights into Brain Connectivity Using Quantitative MRI Measures of White Matter

Andrew L Alexander[1] and Nancy J Lobaugh[2]

[1] Medical Physics and Psychiatry, Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin - Madison
[2] Cognitive Neurology, Sunnybrook Health Sciences Centre, University of Toronto

The vast majority of brain connectivity studies have focused on the activity of measurable brain signals in the cortex and deep gray matter nuclei regions. However, the axons in the white matter serve as the connectivity network of the brain between distant brain regions. Currently, there are not any non-invasive methods for mapping the signal conduction in specific white matter networks. Several MR imaging methods have the potential to provide information related to the physiology and pathology of the white matter tissue substrates, which may ultimately affect brain connectivity.

White matter (WM) is comprised of myelinated axons and glial cells. Axons are the thick branches of neurons, which conduct action potentials (signals) from the neuron cell body to remote target neurons. Myelin is an insulating layer of phospholipids and proteins, which significantly increase the speed of action potential conduction. Either demyelination, myelin degradation, or poor myelin development will impede the efficiency of action potentials and affect neural connectivity. The glia ("brain glue") are non-neural cells and are the supporting cells of the nervous system. They provide support, form myelin, respond to injury, maintain the blood-brain barrier, and regulate the chemical composition of tissue medium. Glial cells include oligodendrocytes (responsible for myelin generation and maintenance), astrocytes (support metabolic function and provide structural support including the blood brain barrier), and microglia (protect the brain from insult and injury). Imaging methods that can characterize the properties of this complex tissue matrix may be valuable for investigating the influence of tissue substrates on neural connectivity.

Conventional MRI is a noninvasive imaging method that can create images with exquisite anatomical detail. While standard MRI methods (e.g., T1-weighted, T2-weighted, proton-density-weighted) can differentiate gray matter and white matter, as well as localize certain brain lesions and abnormalities, it is not quantitative and does not provide information about specific changes in the tissue. However, several quantitative MRI methods have recently been developed which provide either direct or indirect measurements of relevant tissue properties including the microstructural tissue architecture, intra-myelin

water, proteins associated with myelin, axon density, biochemical metabolite concentrations, and response to injury (e.g., inflammation, microglia). These MRI methods include diffusion tensor imaging, magnetization transfer imaging, T1 and T2 relaxometry, MR spectroscopy and spectroscopic imaging, and targeted contrast agents. This chapter will focus on diffusion tensor imaging (DTI), magnetization transfer imaging (MTI) and myelin water fraction imaging (MWFI) using multi-component T2 relaxometry. Although promising, MR spectroscopy is not covered here.

1 Diffusion Tensor Imaging

Diffusion tensor imaging (DTI) is currently the most widely used method for investigations of WM and anatomical connectivity. The diffusion tensor is a simple model of water diffusion in biological tissues and describes the magnitude, anisotropy (directional variation), and orientation of the diffusion distribution.

Diffusion is a random transport phenomenon, which describes the transfer of material (e.g., water molecules) from one spatial location to other locations over time. The Einstein diffusion equation (Einstein 1926):

$$\left\langle \Delta r^2 \right\rangle = 2nD\Delta t \tag{1}$$

states that the mean squared-displacement, $\langle \Delta r^2 \rangle$, from diffusion is proportional to the diffusivity, D (in mm^2/s), over the diffusion time, Δt. The displacement is scaled by the spatial dimensionality, n, which is n = 3 in biological tissues. The diffusivity of pure water at 20°C is roughly $2.0 \times 10{-}3\,mm^2/s$ and slightly higher at body temperature.

The molecules, sub-cellular organelles and cells within biological tissues are in a continuous state of kinetic motion. In particular, water molecules diffuse inside, outside, around, and through cellular structures. The diffusion of water molecules is first caused by random thermal fluctuations. The behavior of the diffusion is further modulated by cytoplasmic currents and the interactions with cellular membranes, and subcellular and organelles.

In fibrous tissues such as white matter tracts in the brain, water diffusion is less hindered or restricted in the direction parallel to the fiber orientation. Conversely, water diffusion is highly restricted or hindered in the directions perpendicular to the fibers. Thus, the diffusion in fibrous tissues is anisotropic. Early diffusion imaging experiments used measurements of parallel ($D_{||}$) and perpendicular ($D_{\perp}$) diffusion components to characterize the diffusion anisotropy (Chenevert et al. 1990; Moseley et al. 1990).

The diffusion tensor is an elegant model of water diffusion (Basser et al. 1994), which assumes that the diffusion is described by a 3D, multivariate normal distribution

$$\mathrm{P}(\Delta\overrightarrow{\mathrm{r}}, \Delta\mathrm{t}) = \frac{1}{\sqrt{(4\pi\Delta\mathrm{t})^3|\mathbf{D}|}} exp\left\{\frac{-\Delta\overrightarrow{\mathrm{r}}^{\mathrm{T}}\mathbf{D}^{-1}\Delta\overrightarrow{\mathrm{r}}}{4\Delta\mathrm{t}}\right\} \tag{2}$$

where Δr is the displacement vector, Δt is the diffusion time, and D is the diffusion tensor, which is a 3×3 matrix

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix}. \tag{3}$$

The diffusion tensor may be diagonalized to calculate the eigenvalues (λ_1, λ_2, λ_3) and corresponding eigenvectors ($\hat{e}_1, \hat{e}_2, \hat{e}_3$) of the diffusion tensor, which describe the relative amplitudes of diffusion and the directions of the principle diffusion axes. A common visual representation of the diffusion tensor is an ellipsoid with the principal axes aligned with the eigenvectors and axes lengths a function of the eigenvalues (see Fig. 1). In the case where the diffusion eigenvalues are (roughly) equal (e.g., $\lambda_1 \sim \lambda_2 \sim \lambda_3$), the diffusion tensor is (nearly) isotropic. When the eigenvalues are significantly different in magnitude (e.g., $\lambda_1 > \lambda_2 > \lambda_3$), the diffusion tensor is anisotropic. Changes in local tissue microstructure with many types of tissue injury, disease or normal physiological changes (i.e., aging) will cause changes in the eigenvalue magnitudes. Thus, the diffusion tensor is an extremely sensitive probe for characterizing both normal and abnormal tissue microstructure.

More specifically in the CNS, water diffusion is typically anisotropic in white matter regions, and isotropic in both gray matter and cerebrospinal fluid (CSF). The major diffusion eigenvector ($\hat{e}_1$-direction of greatest diffusivity) is assumed to be parallel to the tract orientation in regions of homogenous white matter. This directional relationship is the basis for estimating the trajectories of white matter pathways with tractography algorithms.

Diffusion-Weighted Image Acquisition

The random motion of water molecules in biological tissues may cause the signal intensity to decrease in MRI. The NMR signal attenuation from molecular

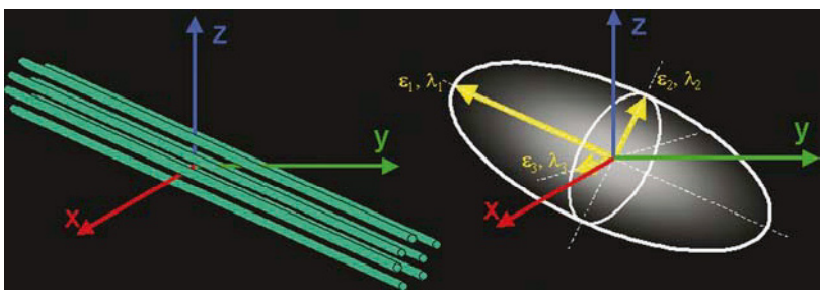


Fig. 1. Schematic representations of diffusion displacement distributions for the diffusion tensor. Ellipsoids (*right*) are used to represent diffusion displacements. The diffusion is highly anisotropic in fibrous tissues such as white matter (*left*). The direction of greatest diffusivity is generally assumed to be parallel to the local direction of white matter

diffusion was first observed more than a half century ago by Hahn (1950). Subsequently, Stejskal & Tanner (1965) described the NMR signal attenuation in the presence of field gradients. More recently, field gradient pulses have been used to create diffusion-weighted MR images (Le Bihan 1990).

Typically, the diffusion weighting is performed using two gradient pulses with equal magnitude and duration (Fig. 2). The first gradient pulse dephases the magnetization across the sample (or voxel in imaging); and the second pulse rephases the magnetization. For stationary (non-diffusing) molecules, the phases induced by both gradient pulses will completely cancel, the magnetization will be maximally coherent, and there will be no signal attenuation from diffusion. In the case of coherent flow in the direction of the applied gradient, the bulk motion will cause the signal phase to change by different amounts for each pulse so that there will be a net phase difference, $\Delta\phi = \gamma v G \delta \Delta$, which is proportional to the velocity, v, the area of the gradient pulses defined by the amplitude, G, and the duration, δ, and the spacing between the pulses, Δ. The gyromagnetic ratio is γ. This is also the basis for phase contrast angiography. For the case of diffusion, the water molecules are also moving, but in arbitrary directions and with variable effective velocities. Thus, in the presence of diffusion gradients, each diffusing molecule will accumulate a different amount of phase. The diffusion-weighted signal is created by summing the magnetization from all water molecules in a voxel. The phase dispersion from diffusion will cause destructive interference, which will cause signal attenuation. For simple isotropic Gaussian diffusion, the signal attenuation for the diffusion gradient pulses in Fig. 2 is described by

$$S = S_o \, e^{-bD} \tag{4}$$

where S is the diffusion-weighted signal, S_o is the signal without any diffusion-weighting gradients (but otherwise identical imaging parameters), D is the apparent diffusion coefficient, and b is the diffusion weighting described by the properties of the pulse pair:

$$b = (\gamma G \delta)^2 (\Delta - \delta/3) \tag{5}$$

Diffusion weighting may be achieved using either a bipolar gradient pulse pair or identical gradient pulses that bracket a 180° refocusing pulse as shown in Fig. 2.

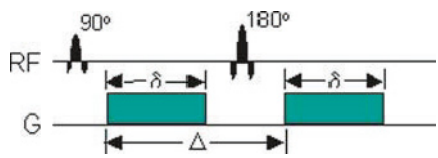


Fig. 2. Spin echo pulse sequence scheme for pulsed-gradient diffusion weighting. A spin-echo refocusing pulse (180°) causes the gradient pulses to be diffusion-weighted

The large gradients make DW MRI extremely sensitive to subject motion. Very small amounts of subject motion may lead to phase inconsistencies in the raw k-space data, causing severe ghosting artifacts in the reconstructed images. Recently, the advances in gradient hardware (maximum gradient amplitude and speed) and the availability of echo planar imaging (EPI) (Mansfield 1984; Turner et al. 1990) on clinical MRI scanners have made routine DW-MRI studies possible. A schematic of a DW-EPI pulse sequence is shown in Fig. 3. With EPI, the image data for a single slice may be collected in 100ms or less, effectively "freezing" any head motion. The fast acquisition speed of EPI makes it highly efficient, which is important for maximizing the image signal-to-noise ratio (SNR) and the accuracy of the diffusion measurements. Thus, single-shot EPI is the most common acquisition method for diffusion-weighted imaging. However, the disadvantages of single shot EPI can also be significant. First, both magnetic field inhomogeneities (Jezzard and Balaban 1995) and eddy currents (Haselgrove and Moore 1996) can warp the image data, thereby compromising the spatial fidelity. Distortions from eddy currents may be either minimized using bipolar diffusion gradient encoding schemes (Alexander et al. 1997; Reese et al. 2003), or corrected retrospectively using image co-registration methods (Haselgrove and Moore 1996; Andersson and Skare 2002; Rohde et al. 2004). Distortions from static field inhomogeneities may be either reduced by using parallel imaging methods such as SENSE (Pruessmann et al. 1999) or retrospectively corrected using maps of the magnetic field (Jezzard and Balaban 1995). Misalignments of k-space data on odd and even lines of k-space will lead to Nyquist or half-field ghosts in the image data. In general, the system should be calibrated to minimize this ghosting although post-processing correction methods have been developed (Zhang and Wehrli 2004). The spatial resolution of 2D EPI pulse sequences also tends

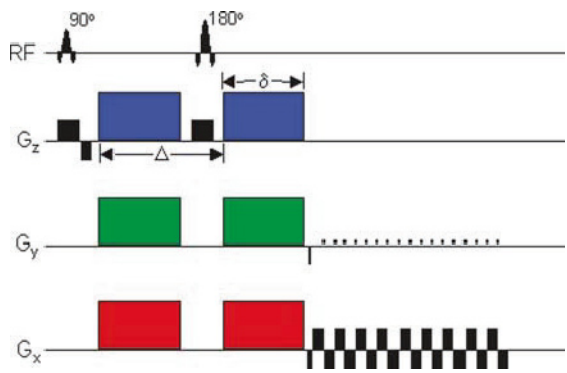


Fig. 3. Schematic of a DW EPI pulse sequence. A spin echo pulse is used to achieve diffusion-weighting from the gradient pulse pairs (colored) as illustrated in Fig. 5. The imaging gradients are shown in black. Diffusion-weighting gradients can be applied in any arbitrary direction using combinations of G_x (*red*), G_y (*green*) and G_z (*blue*)

to be limited. At 1.5T, it is possible to acquire 2.5 mm isotropic voxels over the entire brain in roughly 15 minutes (Jones et al. 2002b). Smaller voxel dimensions may be achieved using either more sensitive RF coils or by going to higher field strengths. Alternative DW imaging techniques, such as PROPELLER (Pipe et al. 2002) and line scan (Gudbjartsson et al. 1997), are less sensitive to motion, eddy currents and B0 distortions.

In the case of anisotropic diffusion, the direction of the diffusion encoding will influence the amount of attenuation. The cartoon in Fig. 4 illustrates the basis for diffusion anisotropy contrast. For anisotropic tissues like white matter, when the diffusion encoding directions are applied parallel to the white matter tract, the signal is highly attenuated. However, when the encoding direction is applied perpendicular to the tract, the diffusion is significantly hindered and the attenuation is much less than in the parallel case. In more isotropic structure regions (such as gray matter), the signal attenuation is independent of the encoding direction.

A minimum of six non-collinear diffusion encoded measurements are necessary to measure the full diffusion tensor (Shrager and Basser 1998; Papadakis et al. 1999). A wide variety of diffusion-tensor encoding strategies with six or more encoding directions have been proposed (e.g., Basser and Pierpaoli 1998; Jones et al. 1999; Papadakis et al. 1999; Shimony et al. 1999; Hasan et al. 2001b). An example of images with DW encoding in twelve directions

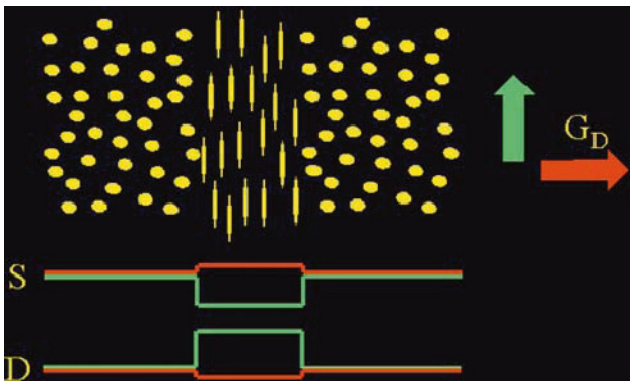


Fig. 4. Illustration of anisotropic signal attenuation with diffusion encoding direction. When the diffusion-weighting (G_D) is applied in the direction parallel (*green*) to the anisotropic cellular structures (e.g., white matter), the signal (S) is strongly attenuated and the apparent diffusivity (D) is high. Conversely, when the diffusion-weighting is applied in the direction perpendicular to the fibrous tissue, the diffusion is less attenuated and the apparent diffusivity is lower. The signal attenuation and diffusivities are independent of the encoding direction in the anisotropic tissue regions. The difference in the directional diffusivities is the source of anisotropy contrast in DTI. The direction of diffusion encoding is selected using different combinations of the diffusion gradients in G_x, G_y and G_z

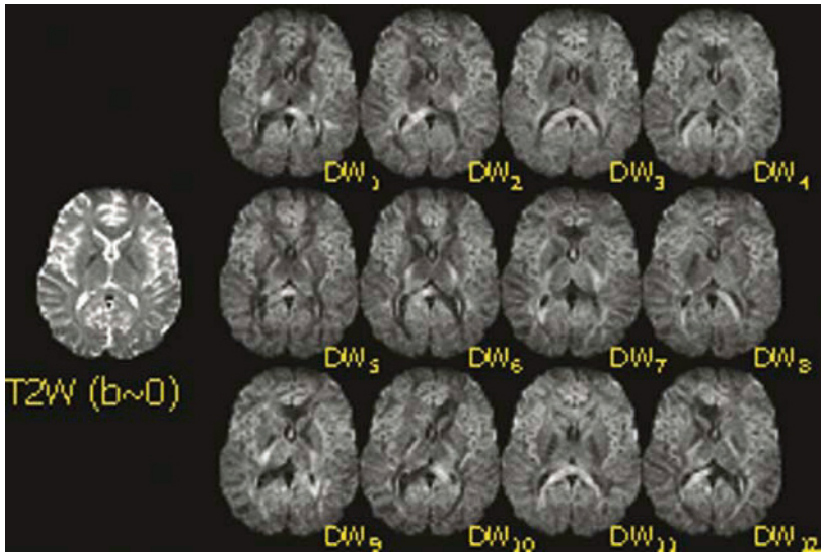


Fig. 5. Example images from a DTI study for a single slice in a human brain. The image on the left is without any diffusion-weighting and is T2-weighted. The twelve images on the right were obtained with diffusion weighting ($b = 1000\,s/mm^2$) applied in twelve non-collinear directions. Note that the image contrast changes significantly with the diffusion encoding direction

for a single slice is shown in Fig. 5. The observed contrast difference for each of the 12 DW encoded images is the basis for the measurement of diffusion anisotropy, which is described later. The selection of tensor encoding directions is critical for accurate and unbiased assessment of diffusion tensor measures. Hasan et al. (2001b) performed a comprehensive comparison of various heuristic, numerically optimized and natural polyhedra encoding sets. This study demonstrated that encoding sets with uniform angular sampling yield the most accurate diffusion tensor estimates. Several recent studies have provided mounting evidence that more diffusion encoding directions causes the measurement errors to be independent of the tensor orientation (e.g., Batchelor et al. 2003; Jones 2004).

There are a number of considerations that should be made when prescribing a diffusion tensor protocol. This is moderately complicated by the wide spectrum of pulse sequence parameters that must be configured. As discussed above, diffusion-weighted, spin-echo, single-shot EPI is the most common pulse sequence for DTI. The optimum diffusion-weighting (also called b-value) for the brain is roughly between 700 and $1300\,s/mm^2$ with a b-value of $1000\,s/mm^2$ being most common. The selection of the number of encoding directions is dependent upon the availability of encoding direction sets, the desired scan time and the maximum number of images that can be obtained in a series. Measurements of diffusion anisotropy tend to be quite sensitive

to image noise, which can also lead to biases in the anisotropy estimates (overestimation of major eigenvalue; underestimation of minor eigenvalue; increase in uncertainty of all eigenvalues) (Pierpaoli and Basser 1996; Chang et al. 2005; Rohde et al. 2005). The accuracy of DTI measures may be improved by either increasing the number of encoding directions or increasing the number of averages. Additional procedures proposed to reduce artifact include the use of peripheral gating to minimize motion related to cardiac pulsitility (Skare and Andersson 2001) and inversion-recovery pulses to minimize partial volume effects from CSF (Bastin 2001; Papadakis et al. 2002; Concha et al. 2005b). Unfortunately, these procedures typically increase the scan time for DTI data collection, and can reduce SNR. The image SNR can also obviously be improved by using larger voxels, although this will increase partial volume averaging of tissues, which can lead to errors in the fits to the diffusion tensor model (Alexander et al. 2001a). The specific parameters for a protocol will depend upon the application. For many routine clinical applications (brain screening, stroke, brain tumors), a fairly coarse spatial resolution can be used with a small number of encoding directions. However, for applications requiring accurate quantification (i.e., quantifying DTI measures in very small white matter tracts, or estimating white matter trajectories with white matter tractography) high spatial resolution is much more important and a large number of diffusion encoding directions or averaging is desirable. High quality DTI data with whole brain coverage, 2.5 mm isotropic resolution and 64 diffusion encoding directions may be obtained in approximately 15 minutes on clinical 1.5T scanners (Jones et al. 2002b). Similar DTI data quality can be achieved in half the time or less at 3.0T, except the image distortions are roughly double.

Diffusion Tensor Image Processing

Maps of DTI measures (mean diffusivity, anisotropy, orientation) are estimated from the raw DW images. As discussed previously, the images may be distorted and misregistered from a combination of eddy currents, subject motion, and magnetic field inhomogeneities. Ideally, these distortions and sources of misregistration should be corrected before calculating any subsequent quantitative diffusion maps. In cases where corrections are not restricted to in-plane errors and distortions, this correction should include recalculation of the diffusion gradient directions or reorienting the tensors (Alexander et al. 2001b; Andersson and Skare 2002; Rohde et al. 2004).

The first step in estimating the diffusion tensor and the associated measures is to calculate the apparent diffusivity maps, $D_{i,app}$, for each encoding direction. The signal attenuation for scalar or isotropic diffusion is described in (4). However, this equation has to be adjusted to describe the signal attenuation for anisotropic diffusion with the diffusion tensor:

$$S_i = S_o \, e^{-b\hat{g}_i^T \mathbf{D} \hat{g}_i} = S_o \, e^{-bD_{i,app}} \tag{6}$$

where S_i is the DW signal in the ith encoding direction, $\hat{g}_i$ is the unit vector describing the DW encoding direction, and b is the amount of diffusion weighting in (6). The apparent diffusivity maps are generated by taking the natural log of (6) and solving for $D_{i,app}$:

$$D_{i,app} = \frac{ln(S_i) - ln(S_o)}{b} \tag{7}$$

This equation works when measurements are obtained for a single diffusion-weighting (b-value) and an image with very little or no diffusion-weighting (S_o). The six independent elements of the diffusion tensor (D_{xx}, D_{yy}, D_{zz}, $D_{xy} = D_{yx}$, $D_{xz} = D_{zx}$, and $D_{yz} = D_{zy}$) may be estimated from the apparent diffusivities using least squares methods (Basser et al. 1994; Hasan et al. 2001a). Maps of the diffusion tensor elements for the data in Fig. 5 are shown in Fig. 6.

Diffusion Tensor Image Measures

The display, meaningful measurement, and interpretation of 3D image data with a 3×3 diffusion matrix at each voxel is a challenging task without simplification of the data. Consequently, it is desirable to distill the image information into simpler scalar maps, particularly for routine clinical applications. The two most common measures are the trace and anisotropy of the

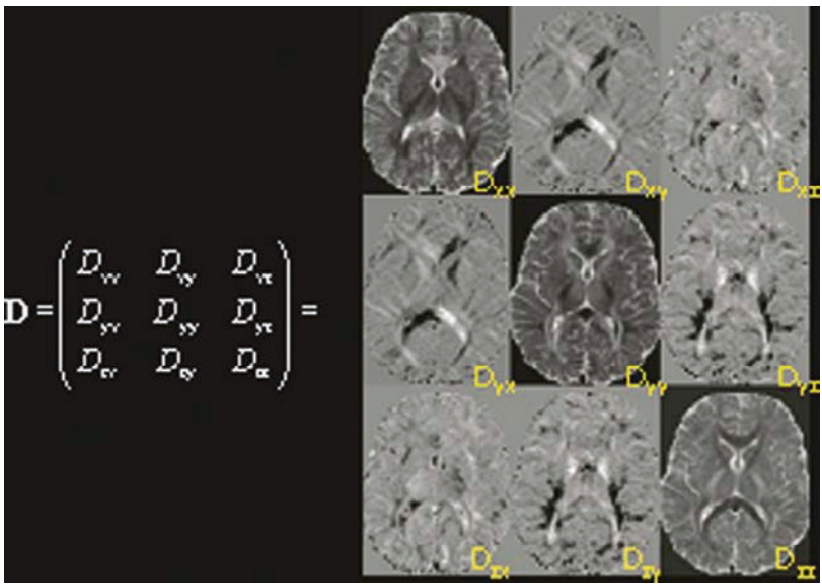


Fig. 6. Maps of the diffusion tensor elements for the DTI data in Fig. 5. Note that the off-diagonal images are symmetric about the diagonal and that the values are both positive and negative

diffusion tensor. The trace of the tensor (Tr), or sum of the diagonal elements of D, is a measure of the magnitude of diffusion and is rotationally invariant. The mean diffusivity, MD, (also called the apparent diffusion coefficient or ADC) is used in many published studies and is simply the trace divided by three (MD = Tr/3). The degree to which the signal is a function of the DW encoding direction is represented by measures of tensor anisotropy. Many measures of anisotropy have been described (Basser and Pierpaoli 1996; Conturo et al. 1996; Pierpaoli and Basser 1996; Ulug and van Zijl 1999; Westin et al. 2002) Most of these measures are rotationally invariant, but do have differential sensitivity to noise (e.g., Skare et al. 2000). Currently, the most widely used invariant measure of anisotropy is the Fractional Anisotropy (FA) described originally by Basser & Pierpaoli (1996).

$$\mathrm{FA} = \sqrt{\frac{(\lambda_1 - \mathrm{MD})^2 + (\lambda_2 - \mathrm{MD})^2 + (\lambda_3 - \mathrm{MD})^2}{2({\lambda_1}^2 + {\lambda_2}^2 + {\lambda_3}^2)}} \tag{8}$$

A third important measure is the tensor orientation described by the major eigenvector direction. For diffusion tensors with high anisotropy, the major eigenvector direction is generally assumed to be parallel to the direction of white matter tract, which is often represented using an RGB (red-green-blue) color map to indicate the eigenvector orientations (Makris et al. 1997; Pajevic and Pierpaoli 1999). The local eigenvector orientations can be used to identify and parcellate specific WM tracts; thus DT-MRI has an excellent potential for applications that require high anatomical specificity. The ability to identify specific white matter tracts on the eigenvector color maps has proven useful for mapping white matter anatomy relative to lesions for preoperative planning (Witwer et al. 2002) and post-operative follow-up (Field et al. 2004). Recently, statistical methods have been developed for quantifying the distributions of tensor orientation in specific brain regions (Wu et al. 2004). Example maps of the mean diffusivity, fractional anisotropy, and major eigenvector direction are shown in Fig. 7.

Relationship to White Matter Physiology & Pathology

The applications of DTI are rapidly growing, in part because the diffusion tensor is exquisitely sensitive to subtle changes or differences in tissue at the microstructural level. DTI studies have found differences in development (e.g., Barnea-Goraly et al. 2005; Snook et al. 2005) and aging (e.g., Abe et al. 2002; Pfefferbaum et al. 2005; Salat et al. 2005), and across a broad spectrum of diseases and disorders including traumatic brain injury (diffuse axonal injury) (Werring et al. 1998; Salmond et al. 2006), epilepsy (Concha et al. 2005a), multiple sclerosis (Cercignani et al. 2000; Rovaris et al. 2002; Assaf et al. 2005), ALS (Ellis et al. 1999; Jacob et al. 2003; Toosy et al. 2003), schizophrenia (Buchsbaum et al. 1998; Lim et al. 1999; Agartz et al. 2001; Jones et al. 2006), bipolar disorder (Adler et al. 2004; Beyer et al. 2005), OCD

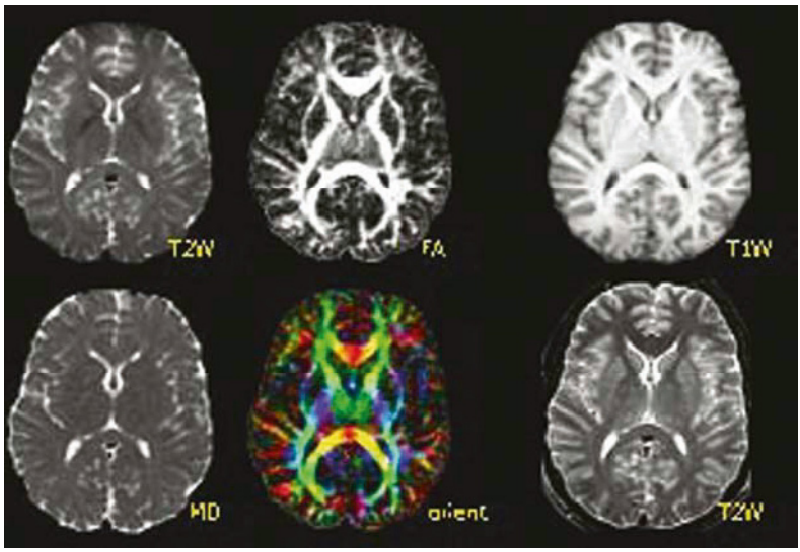


Fig. 7. DTI maps computed from data in Figs. 5 and 6. The images are (*top-left*): T2-weighted "reference" (or b = 0) image from DTI data; (*bottom-left*): mean diffusivity (note similar contrast to T2-W image with CSF appearing hyperintense); (*top-middle*): fractional anisotropy (hyperintense in white matter); (*bottom-middle*) major eigenvector direction indicated by color (red = R/L, green = A/P, blue = S/I) weighted by the FA (note that specific tract groups can be readily identified). Conventional T1-weighted and T2-weighted images (*right column*) at the same anatomical location are shown

(Szeszko et al. 2005), autism (Barnea-Goraly et al. 2004), HIV-AIDs (Pomara et al. 2001; Ragin et al. 2004), and Fragile X (Barnea-Goraly et al. 2003). In nearly all cases, diffusion anisotropy (e.g., fractional anisotropy – FA) is decreased and diffusivity increased in affected regions of diseased white matter relative to healthy controls, while the reverse is true for healthy white matter in development (FA increases, diffusivity decreases).

It is important to note that diffusion anisotropy does not describe the full tensor shape or distribution. This is because different eigenvalue combinations can generate the same values of FA (Alexander et al. 2000). So, for example, a decrease in FA may reflect a decreased major (largest) eigenvalue and/or increased medium/minor (smallest) eigenvalues. FA is likely to be adequate for many applications and appears to be quite sensitive to a broad spectrum of pathological conditions. However, changes simply indicate some difference exists in the tissue microstructure. Several recent studies have looked more directly at the diffusion eigenvalues to determine if they can provide more specific information about the microstructural differences. The results have suggested that the eigenvalue amplitudes or combinations of the eigenvalues (e.g., the radial diffusivity, $D_r = (\lambda_2 + \lambda_3)/2$) demonstrate specific relationships to white matter pathology. For example, the radial diffusivity appears to be specific to myelination in white matter (Song et al. 2005), whereas the

axial diffusivity ($D_a = \lambda_1$) is more specific to axonal density, making it a good model of axonal degeneration (Song et al. 2002). Tensor shape can be fully described using a combination of spherical, linear and planar shape measures (Alexander et al. 2000; Westin et al. 2002), which may also be useful for understanding WM pathology. Consequently, it is important to consider alternative quantitative methods when trying to interpret DTI measurements.

Beyond the Diffusion Tensor

The diffusion tensor is a good model of the diffusion-weighted signal behavior for low levels of diffusion weighting (e.g., $b < 1500\,s/mm^2$). However, the diffusion tensor model does not appear to be consistently accurate in describing the signal behavior for higher levels of diffusion-weighting (e.g., $b > 2000\,s/mm^2$). The problems with the simple diffusion tensor model arise from two sources – (1) apparent "fast" and "slow" diffusing components (Mulkern et al. 1999) that cause the signal decay with diffusion-weighting to appear bi-exponential; and (2) partial volume averaging (e.g.,Alexander et al. 2001a) between tissue groups with distinct diffusion tensor properties (e.g., crossing white matter (WM) tracts, averaging between WM and gray matter tissues). The fast and slow diffusion signals are likely to arise from local restriction effects from cellular membranes although some have hypothesized that these signals correspond to intra- and extra-cellular diffusion.

The effect of partial volume averaging causes ambiguities in the interpretation of diffusion tensor measurements. Whereas the diffusion anisotropy is generally assumed to be high in white matter, regions of crossing white matter tracts will have artifactually low diffusion anisotropy. Consequently, in regions with complex white matter organization, changes or differences in diffusion tensor measures may reflect changes in either the tissue microstructure or the partial volume averaging components. As the diffusion-weighting is increased, the profiles of apparent diffusivity reveal non-Gaussian diffusion behavior in voxels with partial volume averaging.

A growing number of strategies have been developed for measuring and interpreting complex diffusion behavior. The methods vary in their acquisition sampling and analysis approaches. For all of the approaches described here, increasing the maximum diffusion-weighting will improve the characterization of both the slow diffusion components and the partial volume effects, although the measurement SNR will be decreased.

Fast/Slow Diffusion Modeling: Diffusion-weighted measurements over a range of diffusion-weighting have been used to estimate apparent fast and slow components of both apparent diffusivities (BEDI: bi-exponential diffusion imaging) and diffusion tensors (MDTI: multiple diffusion tensor imaging) (Niendorf et al. 1996; Mulkern et al. 1999; Maier et al. 2004). In these cases, the measurements are fit to:

$$S = S_o \left((k)\, e^{-b\hat{g}^T \mathbf{D}_f \hat{g}} + (1-k)\, e^{-b\hat{g}^T \mathbf{D}_s \hat{g}} \right) \tag{9}$$

where $\mathbf{D}_f$ and $\mathbf{D}_s$ are the fast and slow diffusion tensors, and k is the signal fraction from the fast compartment. For a fixed diffusion encoding direction, the signal decay appears bi-exponential with diffusion-weighting. Bi-exponential strategies are appropriate for the cases where there is no significant partial voluming expected and when the diffusion may be modeled using a combination of narrow and broad Gaussian distributions. As discussed earlier, partial volume effects (e.g., crossing WM fibers) will significantly complicate the interpretation of fast and slow diffusing components. In addition, the assignment of these components has been controversial.

High Angular Resolution Diffusion Imaging (HARDI): In order to better characterize the angular diffusion features associated with crossing white matter tracts, several diffusion encoding approaches have been developed that use a large number of encoding directions ($N_e > 40$ up to several hundred) at a fixed level of diffusion-weighting(Alexander et al. 2002; Frank 2002). Although HARDI studies have been reported with diffusion-weighting as low as $b = 1000\,s/mm^2$ (Alexander et al. 2002), the separation of tract components will be much better for higher diffusion-weighting. The original HARDI methods estimated the profiles of apparent diffusion coefficients and used spherical harmonic decomposition methods to estimate the complexity of the diffusion profiles (Alexander et al. 2002; Frank 2002).

Higher order spherical harmonic basis functions represent signal terms that may correspond to crossing white matter tracts in the voxel. Odd spherical harmonic orders do not correspond to meaningful diffusion measurements and are generally assumed to be noise and artifacts.

The HARDI 3D diffusion profiles may also be modeled using generalized diffusion tensor imaging (GDTI) (Ozarslan and Mareci 2003; Liu et al. 2004) which use higher order tensor statistics to model the ADC profile. The GDTI methods proposed by Liu et al. (2004) demonstrate the impressive ability to model asymmetrically bounded diffusion behavior, although the method requires the accurate measurement of the signal phase, which is nearly always discarded and may be difficult to obtain in practice. One problem with these approaches is that in the case of crossing white matter tracts, the directions of maximum ADC do not necessarily correspond to the fiber directions.

One approach to this problem is the q-ball imaging (QBI) solution described by Tuch (2004), which estimates the orientational distribution function (ODF) based upon the Funk-Radon Transform. According to this relationship, the ODF for a particular direction is equivalent to the circular integral about the equator perpendicular to the direction

$$\mathrm{ODF}(\hat{\mathrm{r}}) = \int\int\int_{\mathrm{q}\perp\hat{\mathrm{r}}} \mathrm{E}(\mathbf{q},\Delta)\mathrm{d}^3\mathbf{q} \tag{10}$$

This integral requires that the diffusivities be interpolated over the entire surface of the sphere. Whereas the peaks in the HARDI profile do not necessarily conform to the WM tract directions (see Fig. 8), the peaks in the ODF profiles do in fact correspond to the specific WM tract direction. Since the ODF is

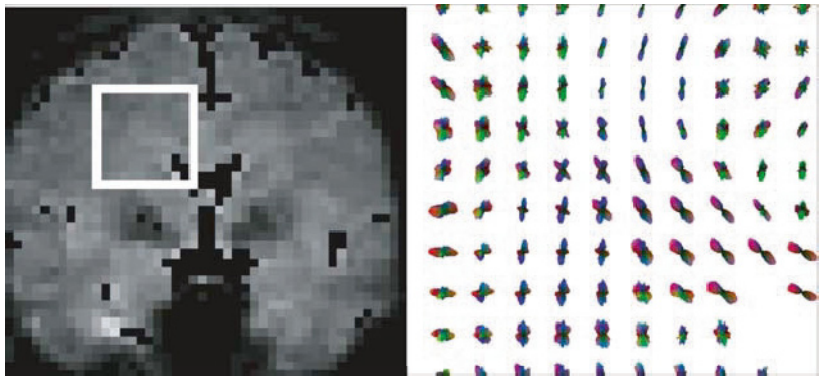


Fig. 8. Example QBI orientational density function (ODF) map for region at the intersection of the corpus callosum, corona radiata and superior longitudinal fasciculus. Regions of crossing WM tracts are clearly observed

estimated by integrating several measurements together, the SNR of the ODF will be much higher than that of the ADC values in the original HARDI.

Diffusion Spectrum Imaging (DSI): The fast/slow diffusion modeling and HARDI approaches represent opposing approaches to complex diffusion characterization. The combination of high angular sampling at multiple levels of diffusion weighting may be used to provide information about both fast/slow diffusion and crossing WM tract orientations. The most basic approach for this application is diffusion spectrum imaging (DSI) (Wedeen et al. 2005) which uses diffusion-weighted samples on a Cartesian q-space lattice, where $\mathbf{q} = \gamma \mathbf{G} \delta$ is the diffusion-weighting wave-vector analogous to wave-vector $\mathbf{k}$ used in k-space sampling for MR image acquisitions. An excellent discussion of q-space imaging is found in the text by Callaghan (1994). For a specified diffusion time, Δ, the probability distribution of diffusion displacements, $\mathrm{P}(\mathbf{R}, \Delta)$, is related to the distribution of sampled diffusion-weighted signals in q-space, $\mathrm{E}(\mathbf{q}, \Delta)$, through a Fourier Transform:

$$\mathrm{P}(\mathbf{R}, \Delta) = \int\int\int \mathrm{E}(\mathbf{q}, \Delta) \mathrm{e}^{-\mathrm{i}2\pi \mathbf{q}.\mathbf{R}} \mathrm{d}^3\mathbf{q} \tag{11}$$

The derivations of q-space formalism assume that the widths of the diffusion-pulses, δ, are narrow relative to the pulse spacing, Δ, such that $\delta << \Delta$. The maximum gradient amplitudes on current clinical MRI systems cause this assumption to be violated for diffusion spectrum imaging, since $\delta \sim \Delta$. The effect of this will be to slightly, but consistently underestimate the diffusion displacements, but the overall distribution shape will be correct (Wedeen et al. 2005). Note that relationship of DSI (q-space) to diffusion tensor imaging is that $\mathrm{P}(\mathbf{R}, \Delta)$ is a multivariate Gaussian and the diffusion-weighting factor is $\mathrm{b} = |\mathrm{q}|^2(\Delta - \delta/3)$ or $\mathrm{b} \sim |\mathrm{q}|^2\Delta$ for small δ. The DSI approach yields empirical estimates of the distributions of diffusion

displacements (e.g., model free), which are described using the standard definitions of Fourier sampling theory.

Since the distributions of diffusion displacements are model independent, the distributions may be challenging to quantify. Several features have been proposed including the zero-displacement probability, $P(\mathbf{R} = 0, \Delta)$, which is higher in regions with more hindered or restricted diffusion; the mean squared displacement,

$$\mathrm{MSD}(\Delta) = \int\int\int P(\mathbf{R}, \Delta)|R|^2 \, d^3\mathbf{R} \tag{12}$$

which is related to the diffusivity (see Fig. 9); the kurtosis of the diffusion distribution, which highlights regions of significant slow diffusion; and the orientational distribution function (ODF)(Wedeen et al. 2005):

$$\mathrm{ODF}(\hat{r}) = \int P(R\hat{r}, \Delta)|R|^2 \, dR \tag{13}$$

Note that this definition of ODF (Eq (9)) for DSI is derived differently for DSI than it is for QBI (Tuch 2004).

While Cartesian sampling facilitates the straightforward FFT for estimation of the displacement densities, Cartesian sampling is not required. Recently, investigators have proposed non-Cartesian sampling strategies of q-space including sampling on concentric spherical shells of constant $|q|$ (Assaf et al. 2004; Wu and Alexander 2005). Assaf et al. then applied a model (CHARMED) of slow and fast diffusing compartments to estimate what they deemed as hindered and restricted diffusion (Assaf et al. 2004). Wu and

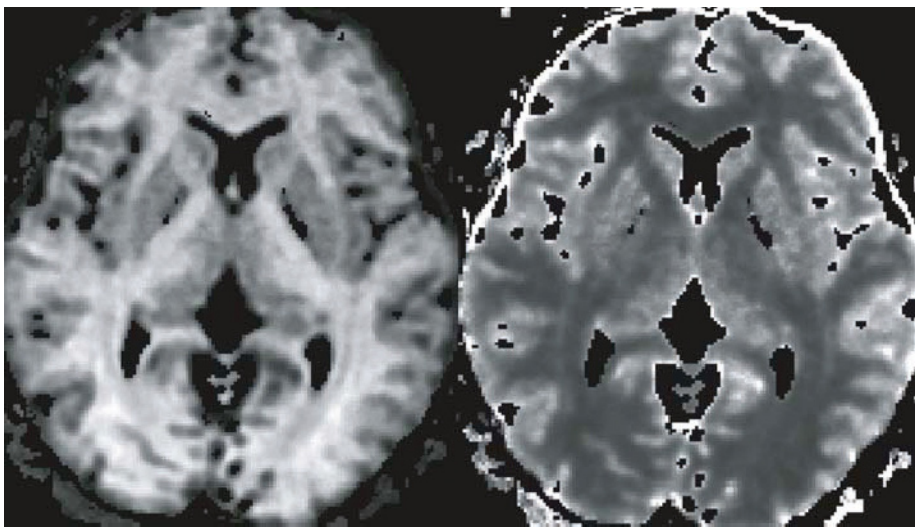


Fig. 9. Example $P(\mathbf{R} = 0;\ \Delta)$ and mean squared displacement maps from DSI study ($N_e = 257;\ \mathrm{bmax} = 9000\,\mathrm{s/mm^2}$)

Alexander (2005) demonstrated that the concentric q-space shell samples in hybrid diffusion imaging (HYDI) could be used for DTI, DSI and QBI in the same experiment.

Applications of High Diffusion-Weighting: The complexity and time required to perform advanced diffusion imaging methods with high diffusion-weighting has limited the number of clinical and research studies relative to the work in diffusion tensor imaging. The pathophysiologic significance of fast/slow diffusion measurements is unclear. Only one published study to date (Brugieres et al. 2004) has specifically examined the effects of pathology (ischemia) on the fast and slow diffusion components. Several small studies of hybrid DSI methods have shown promise in being sensitive to white matter changes associated with multiple sclerosis (Assaf et al. 2002a; Cohen and Assaf 2002), autoimmune neuritis (Assaf et al. 2002b), and vascular dementia (Assaf et al. 2002c). Clearly, more studies are necessary to justify longer imaging times than DTI. To date, none of these methods have been used to directly investigate the relationships to brain connectivity.

From Diffusion to Pathways: White Matter Tractography

In addition to providing information about the mean diffusivity and anisotropy, diffusion imaging methods can also yield novel information about the orientation of local anisotropic tissue features such as bundles of white matter fascicles. In diffusion tensor imaging, the direction of the major eigenvector, $\mathbf{e}_1$, is generally assumed to be parallel to the direction of white matter. This directional information can be visualized by breaking down the major eigenvector into x, y and z components, which can be represented using RGB colors – e.g., Red = e_{1x} =Right/Left; Green= e_{1y} =Anterior/Posterior; Blue= e_{1z} =Inferior/Superior. Maps of WM tract direction can be generated by weighting the RGB color map by an anisotropy measure such as FA (Pajevic and Pierpaoli 1999). For many applications, the use of color labeling is useful for identifying specific WM tracts and visualizing their rough trajectories. An alternative strategy is white matter tractography (WMT), which uses the directional information from diffusion measurements to estimate the trajectories of the white matter pathways. WMT increases the specificity of WM pathway estimates and enables the 3D visualization of these trajectories, which may be challenging using cross-sectional RGB maps.

Deterministic Tractography Algorithms: Most WMT algorithms estimate trajectories from a set of "seed" points. Generally, WMT algorithms may be divided into two classes of algorithms – deterministic (e.g., streamline) and probabilistic (see below). Streamline algorithms are based upon the equation:

$$\mathrm{d}\mathbf{r} = \mathbf{v}_{\mathrm{traj}}\mathrm{d}\tau \tag{14}$$

where $\mathbf{r}(\tau)$ is the path and $\mathbf{v}_{\mathrm{traj}}$ is the vector field that defines the local path direction. Typically, streamline WMT algorithms use major eigenvector field

to define the local trajectory directions $\mathbf{v}_{\text{traj}} = \mathbf{e}_1$ at each step (Conturo et al. 1999; Mori et al. 1999; Basser et al. 2000) (see Fig. 10). Alternatively, tensor deflection (TEND) $\mathbf{v}_{\text{traj}} = \mathbf{D} \cdot \mathbf{v}_{\text{in}}$ uses the entire diffusion tensor to define the local trajectory direction (Lazar et al. 2003). The integration of deterministic pathways may be performed using simple step-wise algorithms including FACT (Mori et al. 1999) and Euler (e.g., $\Delta\mathbf{r} = \mathbf{v}_{\text{traj}}\, \Delta\tau$) (Conturo et al. 1999) integration, or more continuous integration methods such as 2^{nd} or 4^{th} order Runge-Kutta (Basser et al. 2000), which enable more accurate estimates of curved tracts.

Deterministic Tractography Errors: WMT can be visually stunning (see Fig. 11). However, one significant limitation with WMT is that the errors in an estimated tract are generally unknown. Further, the visual aesthetic of WMT, which look like actual white matter patterns, can potentially instill a false sense of confidence in specific results. Unfortunately, there are many potential sources of error that can confound WMT results. Very small perturbations in the image data (i.e., noise, distortion, ghosting, etc.) may lead to significant errors in a complex tensor field such as the brain. Recent studies

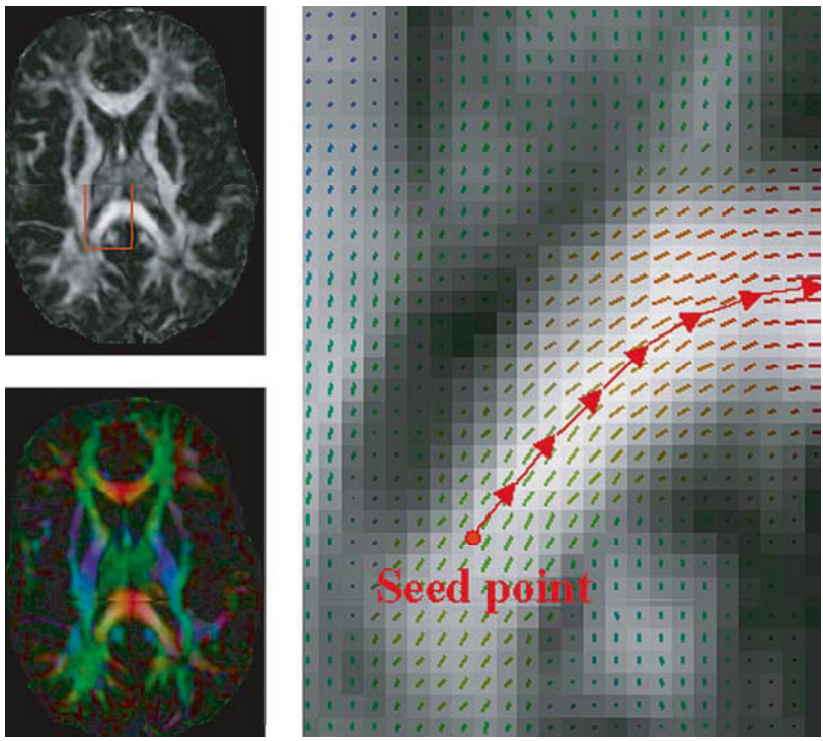


Fig. 10. FA and $\mathbf{e_1}$ color map depicting WM tract orientation. The principle concept of streamline WMT is depicted in a region of corpus callosum. The trajectory is started from a single seed point and the path estimated at discrete steps

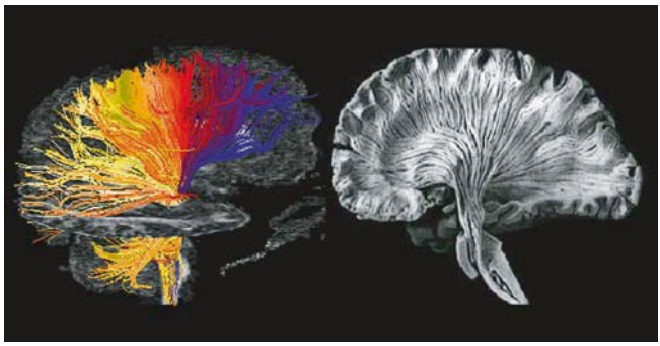


Fig. 11. WMT (*left*) appears to be very similar to an actual white matter dissection (*right*) (Virtual Hospital). http://web.archive.org/web/20050407073533/www.vh.org/adult/provider/anatomy/BrainAnatomy/BrainAnatomy.html

have shown that the dispersion in tract estimates $< \Delta x^2 >$ from image noise is roughly proportional to the distance ($N \cdot w$, where N is the number of voxels and w is voxel size) and inversely proportional to the squares of the eigenvalue differences ($\Delta\lambda_j = \lambda_1 - \lambda_j$) and SNR (Anderson 2001; Lazar et al. 2003)

$$< \Delta x_j{}^2 >= N \cdot w^2 \cdot E/(\Delta\lambda_j \cdot SNR)^2 \tag{15}$$

where E is a factor related to the diffusion tensor encoding scheme and the diffusion tensor orientation, and $j = 2, 3$. Further, the tract dispersion is also affected by the local divergence of the tensor field (Lazar et al. 2003). Even in the complete absence of noise and image artifacts, most current deterministic methods cannot accurately map WM pathways in regions with crossing or converging fibers, which has led to the development of visualization tools to highlight these regions of uncertainty (Jones 2003; Jones et al. 2005c). An alternative approach, recently tested in visual cortex, is likely to be most applicable for mapping interhemispheric fibers. In this method, rather than placing seed voxels in regions of high coherence (e.g., splenium of the corpus callosum), the two hemispheres were seeded separately. Only those obtained tracts that overlapped in the corpus callosum were considered to be valid tracts (Dougherty et al. 2005). This method produced anatomically plausible results for projections from primary visual cortex, but the authors cautioned that many tracts were likely missed, due to the low specificity of WMT and the resolution of current DTI acquisition protocols. New diffusion imaging methods such as DSI and QBI described above are capable of resolving regions of white matter crossing and may ultimately improve WMT in regions of complex WM.

Probabilistic Tractography Algorithms: Although deterministic streamline algorithms are nice tools for visualizing WM patterns, they provide very little information about the reliability of specific results. They rely on accurate placement of seed and deflection point ROIs by the operator, and can vary as a

function of ROI size and shape, making them susceptible to generating highly errant results arising from small errors at a single step. Probabilistic tractography algorithms can overcome some of these limitations. Most probabilistic WMT algorithms are based upon some sort of iterative Monte Carlo approach where multiple trajectories are generated from the seed points with random perturbations to the trajectory directions. Model based tractography algorithms include PICo (Probability Index of Connectivity (Parker et al. 2003), which uses a fast marching technique (Parker et al. 2002), RAVE (Random Vector (Lazar and Alexander 2002)) and ProbTrack (Behrens et al. 2003b). An alternative strategy is to acquire multiple DTI datasets and use bootstrap resampling to derive data-driven estimates of probabilistic tractography (e.g., BOOT-TRAC (Lazar and Alexander 2005) (see Fig. 12). The main difference between model and data-driven approaches is that the variance of the data driven approaches will include the effects of variance in the actual data (e.g., effects of physiologic and artifact noise), not just an idealized model. All of these algorithms create a distribution of tracts, which can be used to estimate the probability of connectivity for the tractography algorithm, which may be used as a surrogate measure of WMT confidence. Additionally, connection probability may be used to segment structures such as the thalamus (Behrens et al. 2003a), cerebral peduncles (Lazar and Alexander 2005), corpus callosum

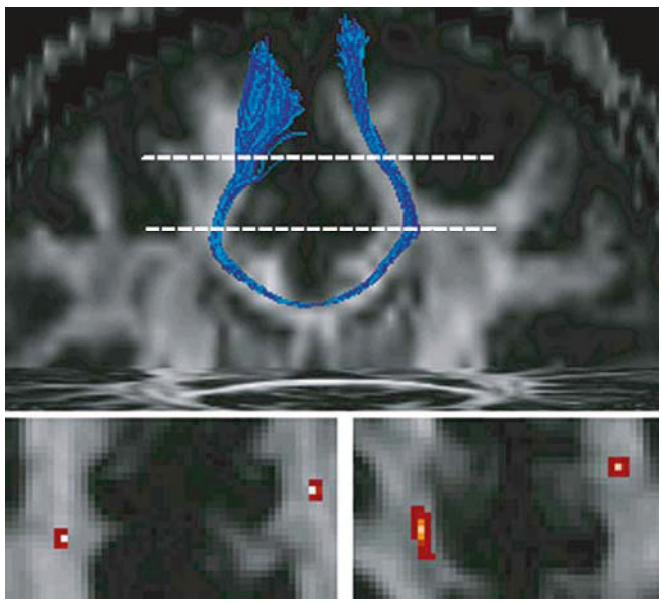


Fig. 12. Probabilistic bootstrap tractography from a single seed point in the corpus callosum illustrating the tract dispersion associated with WMT at two planes above the seed point. The estimated tract density or probability is shown using a hot color scale. The dispersion increases with distance from the seed

(Ciccarelli et al. 2003a), and cortex (Rushworth et al. 2005) according to patterns of maximum connectivity.

Diffusion Imaging and Brain Connectivity: Issues and Considerations

To date, most studies using DTI have focused on analysis of scalar tensor data (anisotropy measures, diffusivity) and have been conducted at three levels of precision: whole-brain histograms; regions-of-interest, and single-voxel analyses. Early studies focused on analysis of whole-brain histograms (e.g., Cercignani et al. 2000; Rovaris et al. 2002), which identify non-specific, global changes in diffusion properties, and may be useful for laying the foundation for more focused analyses. More recently the focus has been on region-of-interest (ROI) and voxel-based analyses. Discussion is ongoing regarding the best methods for accomplishing each type of analysis. When using ROI analyses, it is important to consider the size of the ROI being used, as large ROIs may obscure interesting changes in diffusion measures, and there is a greater possibility that the underlying anatomy will not be consistent across observations. In addition to the usual requirement that the ROIs be placed by a well-trained operator, ROI analyses of DTI data may be may be more sensitive to placement bias in the presence of disease or atrophy. This is especially the case if FA maps are used to define the ROIs. Some have attempted to minimize this potential for bias by lowering the intensity threshold on the FA maps so that local variations in FA are no longer able to guide ROI placement (e.g.,Madden et al. 2004). For voxel-based analyses, the non-diffusion weighted images (b = 0) are often used to register subject data to a common space (Jones et al. 2002a), but this does not guarantee that the underlying fiber architecture (defined by FA or $\hat{e}_1$) is in register. This lack of correspondence is in part due to the high inter-subject variability of the smaller fiber bundles as well as tract characteristics such as their width, neither of which are evident on the b = 0 images. Inter-subject variability is clear when tracts or FA maps are transformed into stereotaxic space. In Fig. 13, optic radiations

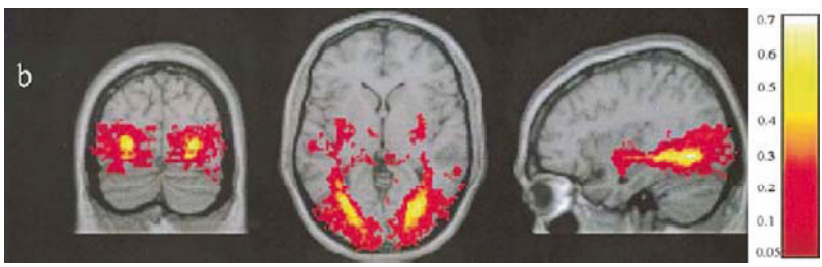


Fig. 13. Optic radiation variability (n = 21). Maximum overlap was 70%. Similar variability would be present if FA maps had been transformed into stereotaxic space. (Reprinted from Ciccarelli et al. 2003b, with permission from Elsevier)

were first identified using probabilistic tractography for individual subjects in native image space. The individual subject data were then resampled into a standardized space, using the b = 0 images as the reference image (Ciccarelli et al. 2003b). Similar dispersion occurs if FA maps are resampled instead of tract probabilities (Jones et al. 2002a).

The large variability across subjects away from tract centers raises the possibility that when correlations of FA and some behavioural or functional measure are found at tissue interfaces, that they may arise simply from the increased variability in FA in these regions. Many published results of voxel-based assessment of group FA differences or FA correlations have identified significant effects in regions of more variable FA. These tend to be located at interfaces of white matter with gray matter or CSF (as seen on T_1-weighted images), or in regions of complex fibre architecture. An example of one such finding is shown in Fig. 14, where correlations of FA with performance on a working-memory task were strongest at tissue interfaces. Because of the error introduced by imperfect registration, residual noise from flow artifact and partial volume effects, as well as the application of smoothing filters (see below), most authors have interpreted such findings with caution. In fact, similar concerns prompted one group to abandon a preliminary voxel-based analysis for one using tractography to define ROIs in the corpus callosum (Kanaan et al. 2006).

Results seem to be more robust to these noise sources if mean tract FA is used rather than voxel-wise FA. An example is seen in recent work examining structure-function relations in the visual system (Toosy et al. 2004). In this study, dispersion was also seen in optic radiations, and it increased as more liberal thresholds were used to define connectivity (Fig. 15, left panel). However, since the regions of high overlap (red) dominated mean FA in the optic radiations, the magnitude of the correlation of FA with the BOLD response in visual cortex was not affected (Fig. 15, right panel).

In voxel-based analyses of functional MRI data, spatial smoothing filters are typically applied to bring the statistical properties of the data more in

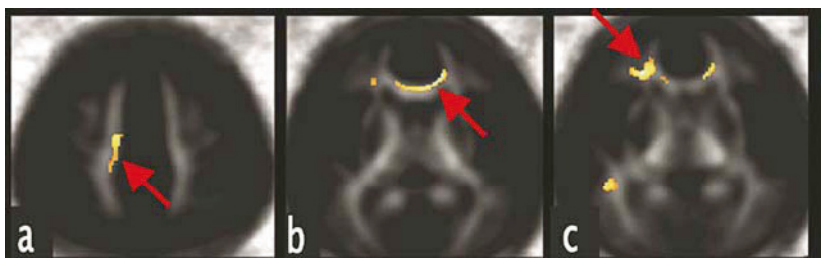


Fig. 14. Example of FA-behavior correlations at tissue interfaces. FA in fronto-parietal white matter (**a**) ranged from 0.2 to 0.6 (n = 21), and correlated with both working memory and BOLD fMRI signal intensity in superior frontal cortex. (Reprinted from Olesen et al. 2003, with permission from Elsevier)

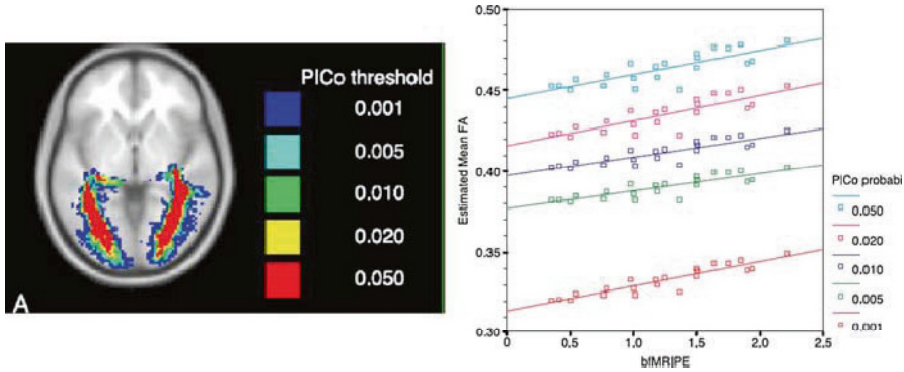


Fig. 15. Left: Optic radiation variability as a function of threshold used to define connectivity (n=22). Right: Mean FA decreased as optic radiation ROI size became larger and more dispersed, but the relation to BOLD response in visual cortex was similar. (Reprinted from Toosy et al. 2004, with permission from Elsevier)

line with random-field theory (Kiebel et al. 1999). It is not yet clear whether smoothing is appropriate for analysis of DTI data, but the size of the smoothing filter can dramatically affect residual errors and the sensitivity to detect group-wise differences (Jones et al. 2005b). In the latter study, significant FA differences between schizophrenic patients and controls were either not found, or were localized to superior temporal sulcus (STS), STS and cerebellum, or cerebellum only. This variability was due only to the size of the smoothing filter, and indicates the reasons for the choice of a specific smoothing filter should be specified.

Alternative methods for voxel-based studies have focused on registering the tensor directly (Xu et al. 2003) or tensor components (Park et al. 2003). Another approach is to use iterative registrations of FA maps to create study-specific templates (Toosy et al. 2004), as is frequently done with voxel-based-morphometry analyses (Good et al. 2001). Finally, a new method has been suggested where non-linear registration is used as the first step in aligning all subjects' FA images together; peak FA "ridges" on are found on the group-averaged FA template, creating a skeleton of the dominant WM tracts. Subject-specific FA values are then derived by finding the location in each subject's data that most closely matches the spatial location of the ridge (Smith et al. 2006, Fig. 16). This approach appears to be robust against residual misregistration since only peak FA values (corresponding to probable tract centers) are analyzed. The use of approaches that attempt to ensure better alignment of tracts across subjects or provide more robust estimates of tract-specific DTI parameters such as FA are critical to furthering our understanding of how alterations in brain connectivity affect brain function and behavior.

Tractography. Obviously, the ability of white matter tractography to estimate patterns of brain connections *in vivo* has piqued the interest of the

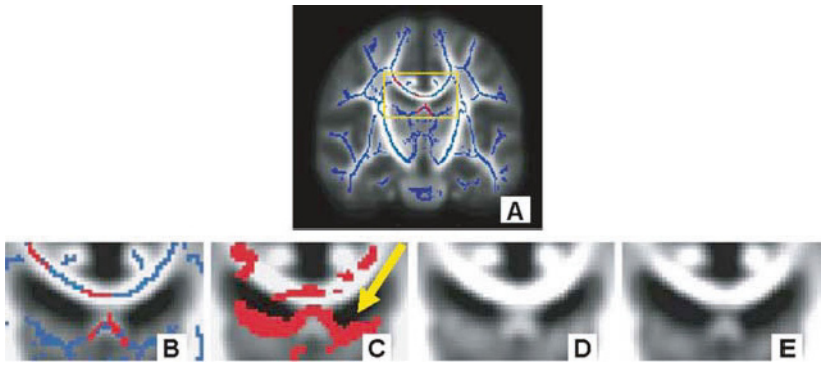


Fig. 16. (**A**) Example of an FA skeleton on a coronal FA map. The outlined region includes the cingulum bundle, corpus callosum, fornix, ventricles and thalamus and is shown in B-E. (**B**) FA skeleton is shown in blue, and significant differences between a group of controls and schizophrenics are in red. (**C**) Voxel-based analysis found additional differences at the lower edge of the ventricles (arrow). (**D**,**E**) Examination of the separate group-mean FA maps indicates this spurious finding was produced because the larger ventricles in the patient group (**E**) were not in register with the controls (**D**). Note that the corpus callosum was well-registered, and the location of FA differences more closely matched the skeletonized FA results. Images courtesy of S. Smith

neuroscience and neuroimaging communities. It is currently the only non-invasive method for reconstructing white matter trajectories in the human brain. Detailed and careful studies using white matter tractography will potentially reveal important information about brain connectivity. However, the links between tractography results, which provide information about anatomical connectivity, and measures of functional and/or effective connectivity (see below) have not yet been clearly established. Several potential anatomical measures that could influence connectivity may be derived from tractography, including the volume, length and/or cross-sectional area of the reconstructed tracts, but these are not routinely applied.

WMT has several potential applications. (1) WMT offers the unique ability to non-invasively visualize the organization of specific WM pathways in individual subjects (e.g., Fig. 11). To date, most studies of white matter neuroanatomy have been conducted using either anatomic dissection methods or axonal tracer studies in animals. The majority of tractography studies have focused on well-known and readily identifiable WM pathways such as the cortico-spinal tract, the corpus callosum and optic radiations. Many of these studies have demonstrated that WMT can generate tract reconstructions that are consistent with known neuroanatomy (e.g., Mori et al. 1999; Stieltjes et al. 2001; Catani et al. 2002; Jellison et al. 2004; Wakana et al. 2004). Recent WMT studies have moved beyond tracking prominent bundles and have attempted to determine the utility of WMT to distinguish between direct and indirect

connections (Catani et al. 2003) and whether highly curved pathways near CSF can be mapped with confidence (Concha et al. 2005b). A common criticism of WMT is that the validation of these results are missing. Two approaches have been applied to address this concern – histopathological measurements and WMT have been compared in animal models (e.g., Burgel et al. 2005; Ronen et al. 2005); and measures of WMT confidence have been developed and applied to provide an estimate of the reliability of specific tractography results. It should also be noted that most neuroimaging results must be interpreted without validation. Thus it is critical to establish the reliability and repeatability of any new WMT method (e.g., Ciccarelli et al. 2003a; Ding et al. 2003; Heiervang et al. 2006). (2) WMT may be used to parcellate specific WM pathways or portions of WM pathways (see Fig. 17). This will enable tract-specific measurements such as tract volume, cross-sectional dimensions, and the statistics of quantitative measurements within the pathways such as mean diffusivity and FA. Several studies have used WMT to perform measurements in specific WM pathways: e.g., fronto-temporal connections in schizophrenia (Jones et al. 2005a; Jones et al. 2006); pyramidal tract development in newborns (Berman et al. 2005), and the pyramidal tracts and corpus callosum in multiple sclerosis (Vaithianathar et al. 2002). Concurrently, progress has been made in the development of tract-specific group templates, which may be useful for voxel-based analyses (Ciccarelli et al. 2003b; Burgel et al. 2005; Johansen-Berg et al. 2005; Thottakara et al. 2006). (3) WMT may be used to visualize specific white matter patterns relative to pathology including brain tumors, M.S. lesions, and vascular malformations. The increased

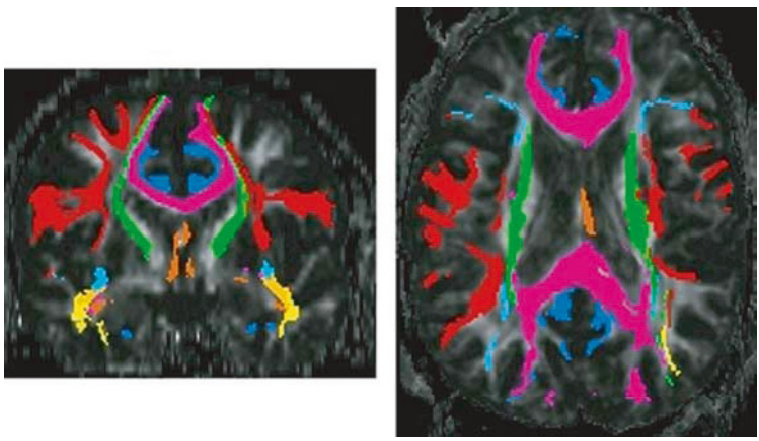


Fig. 17. Parcellation of major white matter pathways using white matter tractography in a single subject. Superior longitudinal fasciculus (*red*); corpus callosum (*purple*); inferior occipital fasciculus (*light blue*); inferior longitudinal fasciculus (*yellow*); uncinate fasciculus (*orange*); fornix/stria terminalis. (*dark orange*); corona radiata (*green*)

specificity of WM trajectories may ultimately be useful for planning surgeries (Holodny et al. 2001; Henry et al. 2004) as well as following the patterns of brain reorganization after surgery (Lazar et al. 2006). However, it should be noted that WMT reconstructions still need further validation before advocating their use as a tool for surgical guidance on a widespread basis. Indeed one recent study demonstrated that their WMT method underestimated the dimensions of the specific tract of interest (Kinoshita et al. 2005). Other studies have started to examine the relationship between specific white matter tracts affected by multiple sclerosis lesions and specific clinical impairments (Lin et al. 2005).

Integrating DTI and WMT with Function

New work is emerging that attempts to do more than simply identify differences in DTI measures as a function of some important variable such as age, disease, or performance. In these studies, the question is: what are the implications of local variations in FA and/or tract characteristics for behavior and brain activity?

Three recent studies examining correlations of local variations in FA with reaction time have found conflicting results. In an ROI analysis, FA was correlated with reaction time in a target-detection task in young and older adults. The results suggested higher FA in the splenium in younger adults and higher FA in the internal capsule in older adults were related to faster reaction times (Madden et al. 2004). Conversely, and somewhat counter intuitively, a voxel-based analysis in a different target detection task revealed primarily positive correlations: high FA was associated with longer reaction times (Tuch et al. 2005), with the strongest effects in the optic radiations. Finally, in traumatic brain injury patients, FA was not correlated with reaction time or cognitive measures, although mean diffusivity did correlate with learning and memory scores (Salmond et al. 2006). Clearly more work is required to understand these relationships.

A more integrative strategy is to examine interactions among FA, BOLD fMRI responses, and behavior or some other external variable, such as age. The few studies attempting to do this have taken a hierarchical approach (e.g., Olesen et al. 2003; Baird et al. 2005). In the first step behavior-FA and behavior-BOLD relations or BOLD activations are assessed separately, effectively reducing the analysis space by creating ROIs from significant clusters. The second step then examines BOLD-FA relations in the smaller subset of regions.

Alternatively, one could ask whether specific tracts are related to behavioural differences. Beaulieu, et al. (2005) used a voxel-based analysis to correlate FA with reading ability in a group of healthy children. The novel aspect to this work was that the authors then used the direction of the principal eigenvector in significant clusters as seeds for WMT. This allowed them to identify potential tracts passing through the significant clusters. They were

able to demonstrate that the largest cluster was more likely associated with a tract not expected to be related to language processing (Fig. 18).

Finally, a number of studies have incorporated diffusion data with the results of fMRI activation studies.. The most common approach has focused on using activated clusters as starting points for tractography to identify anatomical connections. As in any tractography exercise, the choice of which activated voxels to use as seeds for tractography can result in substantially different tracts (Guye et al. 2003). The dependency of tract trajectory on the seed point chosen is compounded by the fact that significant BOLD responses are primarily measured in gray matter, which has generally has low anisotropy, and may be some number of voxels away from highly anisotropic white matter. Since regions of low anisotropy are typically excluded from fibre tracking algorithms, the user must select from nearby voxels with high FA for seeding the WMT. Because of this added uncertainty, it is even more critical to evaluate the robustness of identified tracts. Some progress in tracking between and through gray matter regions has been achieved through the use of probabilistic tractography methods that have been optimized to traverse regions of low anisotropy (e.g., Behrens and Johansen-Berg 2005).

That there is some correspondence between functional and anatomical regions has been recently shown by the Oxford group (Johansen-Berg et al. 2004). In this study, SMA (supplementary motor area) and preSMA were identified in each subject using tasks known to activate those areas independently. Probabilistic tractography was then applied to generate path probabilities from each of the two brain regions. The authors were able to show that separate groups of regions were connected to each of the BOLD regions, with little overlap, as would be expected based on known anatomy. They have recently expanded this analytical approach to show that the functional separation of

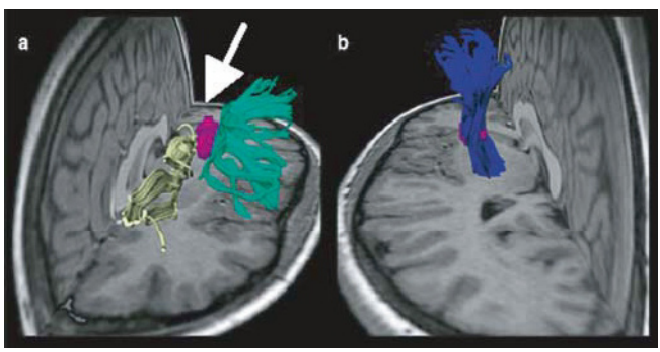


Fig. 18. (**a**) FA in the purple cluster of voxels (*arrow*) correlated with reading ability. Fibre tracking indicated this cluster was in the posterior limb of the internal capsule (**b**), and not in tracts more commonly associated with language (superior longitudinal fasiculus, in green; or superior fronto-occipital fasciculus, in yellow). (Reprinted from Beaulieu et al. 2005, with permission from Elsevier)

these two regions across subjects is more closely aligned to commonalities in local fibre architecture in adjacent white matter than to structural similarities based on conventional T_1-weighted images (Behrens et al. 2006). As the authors point out, they do not yet know if similar relations will hold in other cortical regions. Additionally, the scan time needed to acquire the high resolution DTI dataset (45 min) is not amenable for routine applications. However, the possibility for describing common patterns of functional activations based on common features in the properties of the underlying fibre architecture would be an important adjunct for understanding similarities and differences in brain connectivity.

It is important to keep in mind that DTI tractography is simply defining a model system for brain connectivity. The choice of a particular seed point will influence the derived tracts because of the inherent noise in the data acquisition and the sensitivity of the chosen algorithm to this noise. Tractography is blind to whether the seed point derives from a functional activation or from a well-placed ROI based on expert anatomical knowledge. Therefore, the tracts indicate only the possibility of an anatomical connection between a set of regions; tracts based on functional activations carry no additional "meaning" relative to those derived based on anatomical knowledge. Methods such as those being developed by the Oxford group (e.g., Behrens et al. 2006) will allow for refined anatomical models, but then the task will be to move beyond describing the possibility for information flow to describing how and when information is conveyed along the identified connections.

To fully understand brain function requires more than defining functional "blobs" correlated with some task or behavior. Methods for identifying neural systems and evaluating their interactions have been around for quite some time. Some of the earliest work examined functional connectivity using interregional correlation analyses (e.g., Clark et al. 1984; Horwitz et al. 1984); these were followed with more explicit systems-level analyses of functional and effective connectivity (e.g. Friston et al. 1993; Horwitz et al. 1999; McIntosh 2000), and more recently methods such as dynamic causal modeling (Friston et al. 2003). The importance of moving beyond identifying regions that correlate with some task or behavior has been reemphasized recently by Stephan (2004), who nicely illustrated how two brain regions can correlate independently with a task condition, but have no correlation between themselves (Fig. 19).

The possibilities for incorporating diffusion and other quantitative MRI data into analyses of functional and effective connectivity are many. However it is critical to recognize that simply demonstrating that a pathway exists between two regions that are separately related to some task or behavior does not imply nor guarantee that the identified path mediates the activity between those regions. A more fruitful strategy may be to concurrently determine the existence of pathways between *functionally* connected regions, forming the basis for models of effective connectivity. Regardless of how paths are identified, the information conveyed along those paths should be measured and assessed. Some common and readily available modeling techniques available

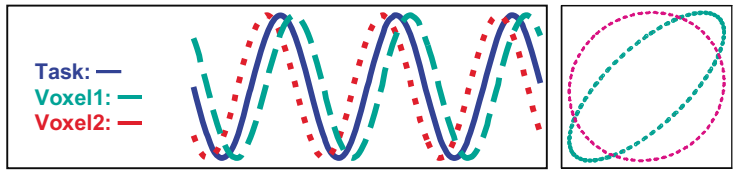


Fig. 19. A) Region A_1 (*red dotted line*) and region A_2 (*green dashed line*) are each correlated with the "task" (*blue, solid line*) at r = 0.73. B) Scatterplot showing that while the correlation of each voxel with the task is high (green, r = 0.73), the correlation between the two voxels is low (magenta, r = 0.07). Adapted from Stephan, 2004

for assessing effective connectivity are reviewed in (McIntosh 2000; Penny et al. 2004; Ramnani et al. 2004; Stephan et al. 2004; Stephan et al. 2005) See also chapters by Bressler and McIntosh, Sporns and Tononi, and Stepan and Friston in this volume. Perhaps the most important contribution from diffusion and other qMRI techniques will come from their ability to provide additional anatomical and physiological constraints to the models. Thus, the confidence that a fibre exists, its length, diameter, "integrity", and myelin content are all important contributions to the regulation of information flow between two regions. Incorporating this information into systems-level analyses of functional imaging data will greatly enhance our understanding of brain function.

Beyond Diffusion

The use of diffusion tensor imaging has become very popular in the last few years, but is not possible to know precisely from DTI studies alone the degree to which observed changes in FA reflect differential changes in myelin composition, fibre number, fibre density or other factors (e.g., Beaulieu 2002; Neil et al. 2002; Prayer and Prayer 2003). Some methods that may help distinguish among these biological properties of white matter are described in the next two sections.

2 Magnetization Transfer

Water provides the largest signal contribution to the MRI signal in brain tissues. While estimates of conductivity can be calculated from diffusion tensor data (Tuch et al. 2001), a more ideal probe of the effectiveness of white matter conduction properties would be obtained from images of myelin components. The problem is that the signals from protein and lipid molecules associated with myelin are essentially undetectable in an MRI experiment because they have ultrashort T2 values (10s of microseconds). However, the magnetization (sum of dipole moments) of free water does interact with the macromolecules

through chemical exchange and dipolar coupling. This exchange of magnetic moments is referred to as magnetization transfer (Balaban and Ceckler 1992).

Magnetization transfer (MT) effects may be detected in an MRI experiment by applying strong radio-frequency (RF) pulses at a frequency shifted by roughly 1000 Hz or more from the resonance frequency of free water. The RF pulse energy will partially saturate the magnetization of the protons bound to macromolecules, which have a very broad frequency spectrum relative to that of free water (width inversely proportional to T2). The fast exchange of magnetization between the macromolecular and free water pools will indirectly attenuate the signal from the free water. The process is illustrated in Fig. 20. The attenuation is a function of the amplitude, rate, and frequency offset of the RF attenuation pulses, and the concentration of macromolecules and exchange rate of the magnetization between the free water and bound macromolecular pools.

The most common approach for characterizing MT is to acquire two sets of images – one with the off-resonance saturation MT pulses (Ms) and one set without (Mo). The MT contrast (MTC) is the difference between the images, MTC = Mo − Ms. Since absolute signal intensities are arbitrary, the MTC is typically normalized by the signal without MT saturation, which is the MT ratio

$$\mathrm{MTR} = (\mathrm{Mo} - \mathrm{Ms})/\mathrm{Mo} \tag{16}$$

The MTR is the most commonly used measure of magnetization transfer and example maps are shown in Fig. 21. Increased MTR values may correspond to increased macromolecular concentrations in the tissue. The MTR values in healthy WM and GM are roughly 0.4-0.55 and 0.25–0.35, respectively. The higher MTR in WM is believed to be associated with the proteins and lipids associated with myelinated axons (Stanisz et al. 1999). Consequently, the MTR in WM is reduced in demyelinating diseases such as multiple sclerosis

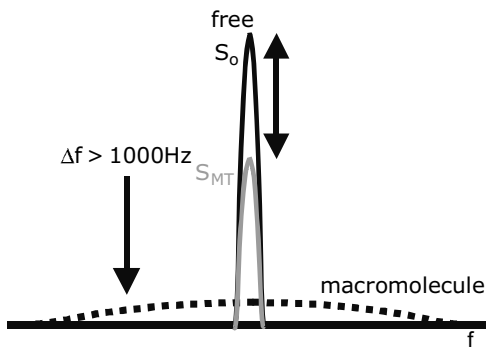


Fig. 20. Schematic of the MT saturation process. An intense RF pulse is applied off-resonance, which saturates the magnetization of the macromolecule pool. Rapid exchange between magnetization of the macromolecule pool and the free water pool causes the free water signal to be partially attenuated

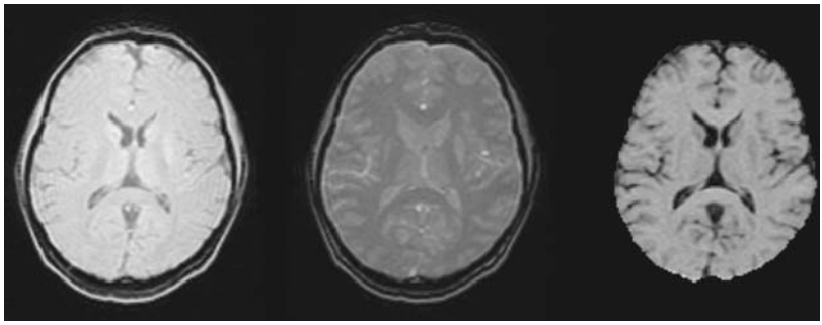


Fig. 21. Example images from an MTR experiment. The image on the left was obtained without any MT saturation. The MT-weighted image in the middle was obtained by applying a 90° pulse 3000 Hz off-resonance (TR = 30 ms). The image on the right is the estimated MT ratio (MTR) map using Equation 16. Images courtesy of A. Samsonov and A. Field

although the MTR can also be influenced by overall water content and other macromolecules in processes such as neuroinflammation (Stanisz et al. 2004).

MT saturation is achieved using an RF pre-pulse, which may be applied in combination with any RF pulse sequence o. An example spin-echo CPMG pulse sequence with RF saturation is shown in Fig. 22. There has been considerable variation of reported MTR properties in the literature, which is likely caused by inconsistencies in the pulse sequence protocols. The exact MTR measurement will depend upon the pulse sequence parameters (e.g., TR, TE, excitation flip angle), the magnetic field strength, as well as the shape, amplitude and frequency offset of the saturation pulses. Consequently, within a single MTR study, the imaging parameters should be fixed to maximize consistency. Common problems with MTR experiments include spatial inhomogeneities in both the static magnetic field (B0) and the RF magnetic field (B1).

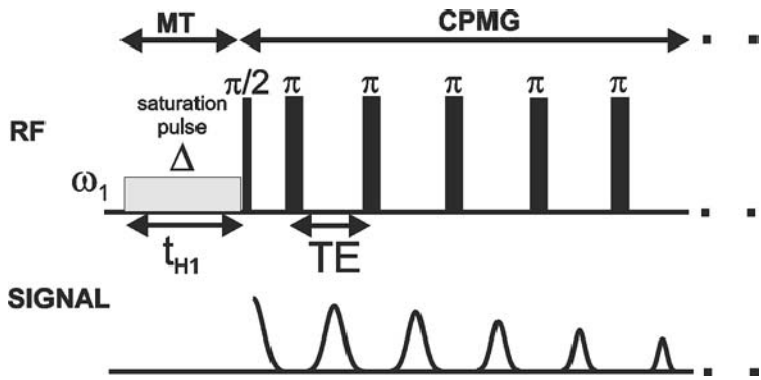


Fig. 22. Measurement of T_2 relaxation in the presence and absence of an RF saturation pulse. Courtesy of G.J. Stanisz

B0 inhomogeneities are caused by poor shimming and spatial variations in the magnetic susceptibilities in soft tissue, bone and air, which lead to shifts (errors) in the saturation frequency offsets. Inhomogeneities in the B1 field, which are common using volume head coils particularly at high magnetic fields ($B0 > 1.5T$) will affect the saturation pulse amplitude and consequently alter the level of MT saturation. Both B0 and B1 fields may be measured and used to retrospectively correct MTR measurements (Sled and Pike 2000; Ropele et al. 2005) Another source of MT saturation is the application of RF excitation pulses for slice selection in 2D pulse sequences (Santyr 1993). The slice selective RF pulses of other slices shifted relative to the current one will cause MT saturation. This is more problematic for multi-slice 2D pulse sequences with many 180° pulses (e.g., fast spin echo, and T1-weighted spin echo); therefore, 3D scans are generally preferable for MTR measurements. Other considerations for MTR measurements are discussed in two excellent review papers (Henkelman et al. 2001; Horsfield et al. 2003).

As discussed above, the MTR measurement is highly dependent upon a broad range of pulse sequence and scanner factors. Consequently, several research groups have been developing models and imaging protocols for quantitative measurements of MT properties (Henkelman et al. 1993; Stanisz et al. 1999; Sled and Pike 2001; Yarnykh 2002; Tozer et al. 2003; Yarnykh and Yuan 2004). these techniques typically require measurements at multiple frequency offsets and saturation pulse amplitudes. Since MT saturation is performed using RF pulses, the MT models are usually based upon a two-pool model (free water and macromolecule) with continuous RF saturation approximated by regular RF saturation pulses. By using these models, it is possible to estimate the macromolecular concentration (bound pool fraction), the exchange rate between the free and bound pools, and the T2 of the bound pool (Fig. 23). Unfortunately, the acquisition of the required images can be quite time consuming, which has limited the overall applicability of the technique. Nonetheless, quantitative MT methods are much more specific than the conventional MTR methods.

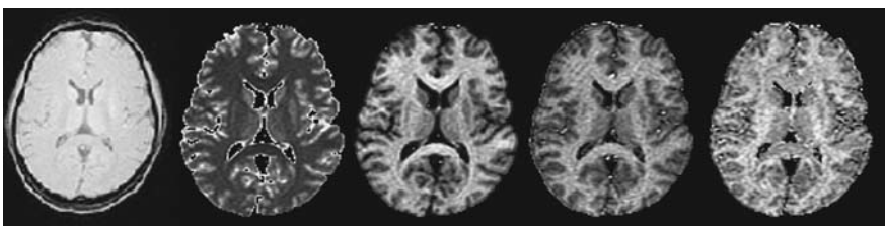


Fig. 23. Quantitative MT maps obtained by acquiring data at multiple frequency offsets and flip angles and using a two pool (free water and macromolecule) model with exchange. The images from left to right are: no MT contrast, T1 map, exchange rate (k), bound pool fraction (f_b), and the T2 of the bound pool ($T2_b$). The images demonstrate the wide range of quantitative imaging measures that can be obtained in a quantitative MT experiment. Images courtesy of A. Samsonov and A. Field

Relationship to Behavioural and Neural Functioning

As for FA, MTR is a non-specific marker of neural damage, such as demyelination. Many of the published MT studies have focused on patients with multiple sclerosis, who show decreased MT in both ROI and whole-brain histogram analyses. In other diseases, results are similar, indicating MTR is a viable marker for affected white and gray matter. MTR has been shown to increase with brain development during the first several years of life (Rademacher et al. 1999; van Buchem et al. 2001) and regional decreases with aging have been found (Armstrong et al. 2004). Differences in MTR were sufficiently large to distinguish patients with mild cognitive impairment from patients with Alzheimer's disease and controls (Kabani et al. 2002a; Kabani et al. 2002b). A number of published studies have also used magnetization transfer methods to compare the brains in patients with schizophrenia against healthy control subjects (Foong et al. 2001; Bagary et al. 2003; Kiefer et al. 2004; Kubicki et al. 2005). Reduced MTR measurements have also been observed in a small sample of patients with late-life major depressive disorders (Kumar et al. 2004).

Only a few studies have attempted to relate magnetization transfer measurements to measures reflecting brain function. A serial MTR study in the optic nerves of 29 patients with acute optic neuritis was performed with measurements of visual system functioning using visual evoked potentials (VEP) (Hickman et al. 2004). No significant differences in MTR were observed between patients and controls at the onset of optic neuritis, although the MTR did decrease in patients over a period of one year. There did not seem to be any direct relationship between MTR and VEP measurements. Another study of 18 patients with early-stage multiple sclerosis (Au Duong et al. 2005) demonstrated a correlation between functional connectivity between left Brodmann areas 45/46 and 24 using an fMRI working memory task, and the MTR of normal appearing white matter and also with brain T2 lesion load. Consequently, the functional connectivity relationship with MTR suggests that changes in the functional working memory network is related to changes in the white matter pathophysiology. A combined MTR and fMRI study (Filippi et al. 2002) of simple motor function in patients with multiple sclerosis revealed correlations between the MTR histogram features of whole-brain, normal appearing brain tissue (both GM and WM) and fMRI signal strengths in ipsilateral sensorimotor cortex and supplementary motor area (bilaterally). The fMRI signal in contralateral sensorimotor cortex was significantly correlated with MTR histogram features in patients with cervical but not dorsal spinal cord myelitis (Rocca et al. 2005). Finally, a recent study measured diffusion and MT in patients with schizophrenia (Kalus et al. 2005). The amygdala showed lower anisotropy (inter-voxel coherence), and differences in quantitative MT measures (T_1, fraction bound pool), but not MTR. The authors interpreted the findings as indicating a possible increase in neuronal density in the amygdala of schizophrenics. The functional significance of these changes is not clear, however, as there were no significant correlations of any of the quantitative MR measures with disease duration or symptom severity.

3 T1 and T2 Relaxometry

Contrast in most human neuroimaging studies is a function of the T1 and T2 relaxation times of the brain tissues. Consequently, regional signal differences in brain images are often caused by differences in the relaxation properties. T1 is the recovery time of the longitudinal magnetization and T2 is the decay constant associated with the transverse magnetization. Both characteristic times are highly sensitive to bulk water of the tissue and tend to increase with water content. Significant changes in both T1 and T2 are observed with early brain maturation (e.g., Miot et al. 1995; Miot-Noirault et al. 1997; Sie et al. 1997; Steen et al. 1997; Paus et al. 2001) and aging (Jernigan et al. 1991; Autti et al. 1994; Salonen et al. 1997). In development, these changes are likely caused by decreased water content and increased water binding and compartmentalization including during premyelination periods when lipids, proteins, and glial cells are increasing. T2 appears to be more sensitive to the changes associated with brain maturation although T1 changes have been reported to be more closely linked to the onset of myelination (e.g., Barkovich et al. 1988; Martin et al. 1988).

There are two principle approaches for measuring T1– inversion recovery and variable saturation. The inversion recovery methods work by inverting the longitudinal magnetization with a 180° pulse and then obtaining measurements with different inversion times. Variable saturation methods work by obtaining measurements with either several RF excitation flip angles or several different TR periods. All methods are highly sensitive to the accuracy of the RF magnetic field, although new analytical methods can retrospectively correct for inhomogeneities (Cheng and Wright 2006).

T2 is generally measured using spin echo pulse sequences, where measurements are obtained at different TE (echo times). The signal decay is governed by the equation S = So exp(-TE/T2). The most efficient method is to use a multiple spin-echo pulse sequence, where measurements are obtained at multiple TE values for a single excitation, although there continue to be lively discussions in the literature concerning the appropriate number and spacing of echos for quantitative T_2 calculations (e.g., Duncan et al. 1996; Whittall et al. 1999; Townsend et al. 2004), related primarily to the nature of T_2 decay (see below). The measurement of T_2 is also highly sensitive to imperfections in the RF and static magnetic fields (Poon and Henkelman 1995). Further, the RF imperfections will also lead to stimulated echoes in multi-echo sequences, which are governed by T1, which can lead to overestimation of the T2. The stimulated echo components can be suppressed using variable amplitude gradient crusher pulses around each 180° refocusing pulse (Poon and Henkelman 1995). As for MT, the accuracy of T2 measurements will depend on these parameters, so if the number of echos possible are limited, they should be chosen with care (Duncan et al. 1996).

In spite of the fact that T1 and T2 are highly sensitive to a wide range of tissue factors, and are therefore likely to be nonspecific, relaxation time

measurements have been shown to be affected in many neurological diseases that have impairments in connectivity including epilepsy, substance abuse and neurodegenerative diseases such as M.S., dementia, schizophrenia, Alzheimer's disease, Parkinson's disease. One potentially confounding factor in many of these studies is the presence of edema, which will increase the bulk water content in the tissue. To date, only one study has specifically related relaxation time measurements to measures of brain connectivity (Vaithianathar et al. 2002). In this study of MS patients, DTI was used to identify the pyramidal tracts and fibers passing through the corpus callosum. Histograms of T_1 relaxation data along the pathways were generated and indicated decreased T_1 relaxivity in patients relative to controls. There was no correlation of T_1 relaxation in these paths with standard clinical disability rating scale scores, and no cognitive measures were available for analysis.

Myelin Water Fraction

Although the specificity of T1 and T2 measurements are generally perceived as being poor, several investigators have recently shown that the T2 signal decay in neural tissue is multi-exponential with echo time (Menon and Allen 1991; MacKay et al. 1994; Whittall et al. 1997). Further investigation has shown that different water tissue compartments each have distinct T2 characteristics, and may be separated (see Fig. 24). In white matter, the water signal compartments are believed to originate from components of free water (e.g., edema, CSF, which have long T2 $>$ 120 ms), extracellular water (T2 $\sim$ 60–90 ms) and water within the myelin membranes of axons (T2 $\sim$ 10–40 ms) (MacKay et al. 1994; Beaulieu et al. 1998; Stanisz and Henkelman 1998; Vavasour et al. 1998; Laule et al. 2004). The T_2 of the extracellular fraction can be used to identify inflamed neural tissues (Stanisz et al. 2004), and the latter component is of significant interest because it is specific to myelin, which is critical for signal conduction in the brain. Consequentially, a potentially important biomarker is the myelin water fraction, which is the total signal

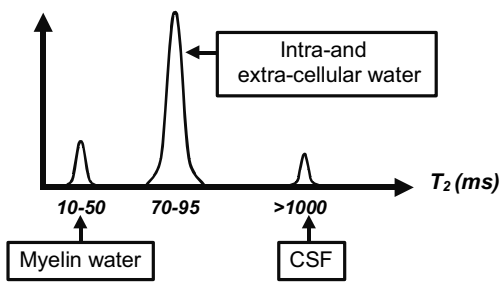


Fig. 24. T2 spectrum of water signal in white matter. The water in the myelin layers has a very short T2 (between 10 and 50 ms), intra- and extra-cellular water have intermediate T2 values, and CSF and unrestricted water pools have much longer T2 values

from the short T2 signal component relative to the total signal from both the short and intermediate tissue signal components. In healthy adult WM, the myelin water fraction (MWF) is typically 6–15% dependent upon the region (Whittall et al. 1997). A representative map of MWF is shown in Fig. 25.

Measurements of MWF are usually obtained using a 2D multiple spin echo sequence, which consists of a train of equally spaced 180° refocusing pulses (Poon and Henkelman 1992; Poon and Henkelman 1995; Whittall et al. 1997). T2 measurements are highly sensitive to errors in the RF magnetic field, which are problematic for typical slice-selective RF refocusing pulses. Consequently, non-selective refocusing pulses are often used, which limits the acquisition to a single 2D slice. Variable amplitude crusher gradient pulses are typically placed around each refocusing pulse to suppress the signal from stimulated echoes. The fitting of the T2 model is also highly sensitive to image noise; consequently, long scan times are typically required to achieve sufficient SNR. Different strategies exist for fitting the T2 signal decay to a multi-exponential function (e.g., Stanisz and Henkelman 1998; Webb et al. 2003; Jones et al. 2004) although the non-negative least squares (NNLS) method is probably most commonly used (Whittall et al. 1997). The slow acquisition time (typically $>$ 10 minutes) for a single 2D slice has ultimately limited the application of this approach. However, one consideration is that the 2D imaging times are in line with MR spectroscopy. Further, the MWF is one of the more specific measures of white matter tissue properties, which makes it promising for correlations with measures of brain connectivity. Careful selection of echos in conventional pulse sequences may provide reasonable myelin maps (Vidarsson et al. 2005; Oh et al. 2006), although the option to acquire such data is not available routinely on most clinical scanners. Future

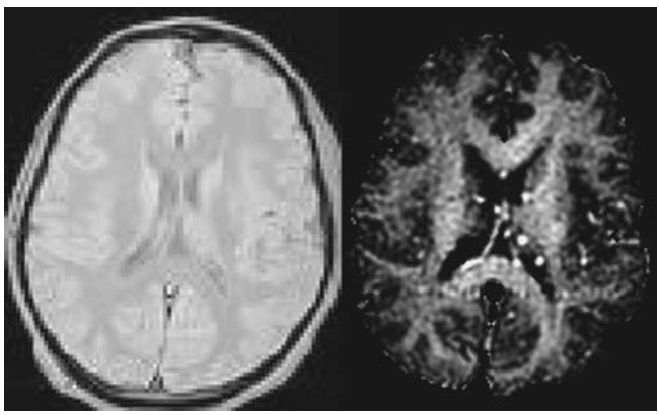


Fig. 25. Maps from a myelin water fraction experiment. The image on the left is a proton-density weighted image obtained from the first TE (8 ms) in the CPMG echo train. The map on the right is the estimated myelin water fraction image at the same slice location. Note that the myelin water fraction is much higher in regions of white matter

developments are clearly needed to improve both the acquisition speed and spatial coverage of the technique, which are somewhat at odds with one another. Imaging at higher magnetic field strengths, with better RF coils, parallel imaging and 3D pulse sequences may ultimately improve the utility of the method.

To date, no studies have been performed which have related MWF measurements to measures of brain connectivity. However, MWF measurements in WM have been shown to be affected in brain diseases with aberrant brain connectivity behavior including schizophrenia (Flynn et al. 2003) and multiple sclerosis (Vavasour et al. 1998; Gareau et al. 2000; Whittall et al. 2002; Laule et al. 2004; Tozer et al. 2005).

4 Discussion and Future Directions

In this chapter, we have described several quantitative MRI measures, which are promising for the characterization of brain connectivity. However, to date, there has been a relative paucity of experiments that have directly compared functional and effective measures of brain connectivity with these structural and anatomical measures of brain connectivity and physiology. This is likely to change in the near future as these techniques become more available. Characterization of WM anatomy and physiology with MRI may enable more complex models of brain connectivity to be developed, as the circuitry of brain connectivity becomes more well-defined. For example, many have proposed that FA increases are primarily reflecting myelination. This leads to the prediction that FA would be correlated most strongly with the short myelin-water fraction from T_2 relaxometry experiments, as well as to the size of the macromolecular pool in quantitative MT studies. On the other hand, if changes in fibre density underlie changes in FA (Beaulieu 2002), FA should be more strongly associated with the extracellular water peak. Preliminary evidence for this prediction comes from recent work showing FA was not correlated with the myelin water fraction in white matter (MacKay et al. 2006). Functional and effective connectivity studies so far have generally modeled the brain as a "black box" with inputs and outputs, and most of the internal circuitry has been derived from non-human primate studies. Quantitative structural and physiological image data from MRI may provide critical information about the functional circuitry within the black box.

To move quantitative MRI into the forefront of techniques for characterizing brain connectivity, further developments are necessary. Obviously, improvements to both the imaging technology through better and more efficient pulse sequences, imaging RF coils, and gradient coils, and quantitative imaging models and image analysis methods will facilitate comparisons between more conventional connectivity measures with quantitative MRI measures of WM. However, even with improvements in the technology, the application will be somewhat limited unless they become more readily available, either

through the MRI system manufacturers or through research collaborations. While the methodologies are still young and emerging, we can already pose some interesting questions: Do variations in diffusion parameters or myelin content along tracts relate to function? Is the whole fibre tract affected? Does knowing something about tract likelihood help predict differences in functional and effective connectivity? Answers to these and similar questions will require multimodal imaging, as most quantitative MRI studies have focused on a single measure or measurement type We will also need a better understanding of the statistical properties of the data, and sophisticated multivariate and nonlinear modeling techniques, some of which are already available, and others of which are discussed throughout this volume. This will be an iterative process and will require refinement of both imaging and analysis techniques. However, we have optimism that in the end the model fits will be acceptable and we will know something useful about how brain structure contributes to brain function.

Acknowledgements

The authors would like to thank Mariana Lazar, Yu-Chien Wu, Alexey Samsonov, Yijing Wu, and Aaron Field at the University of Wisconsin for providing figures, as well as Steve Smith from Oxford and Greg Stanisz from Sunnybrook Health Sciences Centre, and thanks also to Ana Solodkin at the University of Chicago for lively discussions. Some of the diffusion MRI work presented here was supported by NIH grant RO1 MH 62015 and the MT and MWF data was generated from a grant from the National Multiple Sclerosis Society (University of Wisconsin).

References

Abe O, Aoki S, Hayashi N, Yamada H, Kunimatsu A, Mori H, Yoshikawa T, Okubo T, Ohtomo K (2002) Normal aging in the central nervous system: quantitative MR diffusion-tensor analysis. Neurobiology of Aging 23: 433–441

Adler CM, Holland SK, Schmithorst V, Wilke M, Weiss KL, Pan H, Strakowski SM (2004) Abnormal frontal white matter tracts in bipolar disorder: a diffusion tensor imaging study. Bipolar Disorders 6: 197–203

Agartz I, Andersson JL, Skare S (2001) Abnormal brain white matter in schizophrenia: a diffusion tensor imaging study. Neuroreport 12: 2251–2254

Alexander AL, Hasan K, Kindlmann G, Parker DL, Tsuruda JS (2000) A geometric analysis of diffusion tensor measurements of the human brain. Magnetic Resonance in Medicine 44: 283–291

Alexander AL, Hasan KM, Lazar M, Tsuruda JS, Parker DL (2001a) Analysis of partial volume effects in diffusion-tensor MRI. Magnetic Resonance in Medicine 45: 770–780

Alexander AL, Tsuruda JS, Parker DL (1997) Elimination of eddy current artifacts in diffusion-weighted echo-planar images: the use of bipolar gradients. Magnetic Resonance in Medicine 38: 1016–1021

Alexander DC, Barker GJ, Arridge SR (2002) Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data. Magnetic Resonance in Medicine 48: 331–340

Alexander DC, Pierpaoli C, Basser PJ, Gee JC (2001b) Spatial transformations of diffusion tensor magnetic resonance images. IEEE Transactions on Medical Imaging 20: 1131–1139

Anderson AW (2001) Theoretical analysis of the effects of noise on diffusion tensor imaging. Magnetic Resonance in Medicine 46: 1174–1188

Andersson JL, Skare S (2002) A model-based method for retrospective correction of geometric distortions in diffusion-weighted EPI. Neuroimage 16: 177–199

Armstrong CL, Traipe E, Hunter JV, Haselgrove JC, Ledakis GE, Tallent EM, Shera D, van Buchem MA (2004) Age-related, regional, hemispheric, and medial-lateral differences in myelin integrity in vivo in the normal adult brain. AJNR: American Journal of Neuroradiology 25: 977–984

Assaf Y, Ben-Bashat D, Chapman J, Peled S, Biton IE, Kafri M, Segev Y, Hendler T, Korczyn AD, Graif M, Cohen Y (2002a) High b-value q-space analyzed diffusion-weighted MRI: application to multiple sclerosis. Magnetic Resonance in Medicine 47: 115–126

Assaf Y, Chapman J, Ben-Bashat D, Hendler T, Segev Y, Korczyn AD, Graif M, Cohen Y (2005) White matter changes in multiple sclerosis: correlation of q-space diffusion MRI and 1H MRS. Magnetic Resonance Imaging 23: 703–710

Assaf Y, Freidlin RZ, Rohde GK, Basser PJ (2004) New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter. Magnetic Resonance in Medicine 52: 965–978

Assaf Y, Kafri M, Shinar H, Chapman J, Korczyn AD, Navon G, Cohen Y (2002b) Changes in axonal morphology in experimental autoimmune neuritis as studied by high b-value q-space (1)H and (2)H DQF diffusion magnetic resonance spectroscopy. Magnetic Resonance in Medicine 48: 71–81

Assaf Y, Mayzel-Oreg O, Gigi A, Ben-Bashat D, Mordohovitch M, Verchovsky R, Reider G, II, Hendler T, Graif M, Cohen Y, Korczyn AD (2002c) High b value q-space-analyzed diffusion MRI in vascular dementia: a preliminary study. Journal of the Neurological Sciences 203–204: 235–239

Au Duong MV, Audoin B, Boulanouar K, Ibarrola D, Malikova I, Confort-Gouny S, Celsis P, Pelletier J, Cozzone PJ, Ranjeva JP (2005) Altered functional connectivity related to white matter changes inside the working memory network at the very early stage of MS. Journal of Cerebral Blood Flow and Metabolism 25: 1245–1253

Autti T, Raininko R, Vanhanen SL, Kallio M, Santavuori P (1994) MRI of the normal brain from early childhood to middle age. II. Age dependence of signal intensity changes on T2-weighted images. Neuroradiology 36: 649–651

Bagary MS, Symms MR, Barker GJ, Mutsatsa SH, Joyce EM, Ron MA (2003) Gray and white matter brain abnormalities in first-episode schizophrenia inferred from magnetization transfer imaging. Archives of General Psychiatry 60: 779–788

Baird AA, Colvin MK, Vanhorn JD, Inati S, Gazzaniga MS (2005) Functional connectivity: integrating behavioral, diffusion tensor imaging, and functional magnetic resonance imaging data sets. Journal of Cognitive Neuroscience 17: 687–693

Balaban RS, Ceckler TL (1992) Magnetization transfer contrast in magnetic resonance imaging. Magnetic Resonance Quarterly 8: 116–137

Barkovich AJ, Kjos BO, Jackson DE, Jr., Norman D (1988) Normal maturation of the neonatal and infant brain: MR imaging at 1.5 T. Radiology 166: 173–180

Barnea-Goraly N, Eliez S, Hedeus M, Menon V, White CD, Moseley M, Reiss AL (2003) White matter tract alterations in fragile X syndrome: preliminary evidence from diffusion tensor imaging. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics 118: 81–88

Barnea-Goraly N, Kwon H, Menon V, Eliez S, Lotspeich L, Reiss AL (2004) White matter structure in autism: preliminary evidence from diffusion tensor imaging. Biological Psychiatry 55: 323–326

Barnea-Goraly N, Menon V, Eckert M, Tamm L, Bammer R, Karchemskiy A, Dant CC, Reiss AL (2005) White matter development during childhood and adolescence: A cross-sectional diffusion tensor imaging study. Cerebral Cortex 15: 1848–1854

Basser PJ, Mattiello J, LeBihan D (1994) Estimation of the effective self-diffusion tensor from the NMR spin echo. Journal of Magnetic Resonance Series B 103: 247–254

Basser PJ, Pajevic S, Pierpaoli C, Duda J, Aldroubi A (2000) In vivo fiber tractography using DT-MRI data. Magnetic Resonance in Medicine 44: 625–632

Basser PJ, Pierpaoli C (1996) Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. Journal of Magnetic Resonance Series B 111: 209–219

Basser PJ, Pierpaoli C (1998) A simplified method to measure the diffusion tensor from seven MR images. Magnetic Resonance in Medicine 39: 928–934

Bastin ME (2001) On the use of the FLAIR technique to improve the correction of eddy current induced artefacts in MR diffusion tensor imaging. Magnetic Resonance Imaging 19: 937–950

Batchelor PG, Atkinson D, Hill DL, Calamante F, Connelly A (2003) Anisotropic noise propagation in diffusion tensor MRI sampling schemes. Magnetic Resonance in Medicine 49: 1143–1151

Beaulieu C (2002) The basis of anisotropic water diffusion in the nervous system - a technical review. NMR in Biomedicine 15: 435–455

Beaulieu C, Fenrich FR, Allen PS (1998) Multicomponent water proton transverse relaxation and T2-discriminated water diffusion in myelinated and nonmyelinated nerve. Magnetic Resonance Imaging 16: 1201–1210

Beaulieu C, Plewes C, Paulson LA, Roy D, Snook L, Concha L, Phillips L (2005) Imaging brain connectivity in children with diverse reading ability. Neuroimage 25: 1266–1271

Behrens TE, Jenkinson M, Robson MD, Smith SM, Johansen-Berg H (2006) A consistent relationship between local white matter architecture and functional specialisation in medial frontal cortex. Neuroimage 30: 220–227

Behrens TE, Johansen-Berg H (2005) Relating connectional architecture to grey matter function using diffusion imaging. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 360: 903–911

Behrens TE, Johansen-Berg H, Woolrich MW, Smith SM, Wheeler-Kingshott CA, Boulby PA, Barker GJ, Sillery EL, Sheehan K, Ciccarelli O, Thompson AJ, Brady JM, Matthews PM (2003a) Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nature Neuroscience 6: 750–757

Behrens TE, Woolrich MW, Jenkinson M, Johansen-Berg H, Nunes RG, Clare S, Matthews PM, Brady JM, Smith SM (2003b) Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magnetic Resonance in Medicine 50: 1077–1088

Berman JI, Mukherjee P, Partridge SC, Miller SP, Ferriero DM, Barkovich AJ, Vigneron DB, Henry RG (2005) Quantitative diffusion tensor MRI fiber tractography of sensorimotor white matter development in premature infants. Neuroimage 27: 862–871

Beyer JL, Taylor WD, MacFall JR, Kuchibhatla M, Payne ME, Provenzale JM, Cassidy F, Krishnan KR (2005) Cortical white matter microstructural abnormalities in bipolar disorder. Neuropsychopharmacology 30: 2225–2229

Brugieres P, Thomas P, Maraval A, Hosseini H, Combes C, Chafiq A, Ruel L, Breil S, Peschanski M, Gaston A (2004) Water diffusion compartmentation at high b values in ischemic human brain. AJNR: American Journal of Neuroradiology 25: 692–698

Buchsbaum MS, Tang CY, Peled S, Gudbjartsson H, Lu D, Hazlett EA, Downhill J, Haznedar M, Fallon JH, Atlas SW (1998) MRI white matter diffusion anisotropy and PET metabolic rate in schizophrenia. Neuroreport 9: 425–430

Burgel U, Amunts K, Hoemke L, Mohlberg H, Gilsbach JM, Zilles K (2005) White matter fiber tracts of the human brain: Three-dimensional mapping at microscopic resolution, topography and intersubject variability. Neuroimage

Callaghan PT (1994) Principles of Magnetic Resonance Microscopy Oxford University Press, Oxford

Catani M, Howard RJ, Pajevic S, Jones DK (2002) Virtual in vivo interactive dissection of white matter fasciculi in the human brain. Neuroimage 17: 77–94

Catani M, Jones DK, Donato R, Ffytche DH (2003) Occipito-temporal connections in the human brain. Brain 126: 2093–2107

Cercignani M, Iannucci G, Rocca MA, Comi G, Horsfield MA, Filippi M (2000) Pathologic damage in MS assessed by diffusion-weighted and magnetization transfer MRI. Neurology 54: 1139–1144

Chang LC, Jones DK, Pierpaoli C (2005) RESTORE: robust estimation of tensors by outlier rejection. Magnetic Resonance in Medicine 53: 1088–1095

Chenevert TL, Brunberg JA, Pipe JG (1990) Anisotropic diffusion in human white matter: demonstration with MR techniques in vivo. Radiology 177: 401–405

Cheng HL, Wright GA (2006) Rapid high-resolution T(1) mapping by variable flip angles: accurate and precise measurements in the presence of radiofrequency field inhomogeneity. Magnetic Resonance in Medicine 55: 566–574

Ciccarelli O, Parker GJ, Toosy AT, Wheeler-Kingshott CA, Barker GJ, Boulby PA, Miller DH, Thompson AJ (2003a) From diffusion tractography to quantitative white matter tract measures: a reproducibility study. Neuroimage 18: 348–359

Ciccarelli O, Toosy AT, Parker GJ, Wheeler-Kingshott CA, Barker GJ, Miller DH, Thompson AJ (2003b) Diffusion tractography based group mapping of major white-matter pathways in the human brain. Neuroimage 19: 1545–1555

Clark CM, Kessler R, Buchsbaum MS, Margolin RA, Holcomb HH (1984) Correlational methods for determining regional coupling of cerebral glucose metabolism. A pilot study. Biological Psychiatry 19: 663–678

Cohen Y, Assaf Y (2002) High b-value q-space analyzed diffusion-weighted MRS and MRI in neuronal tissues - a technical review. NMR in Biomedicine 15: 516–542

Concha L, Beaulieu C, Gross DW (2005a) Bilateral limbic diffusion abnormalities in unilateral temporal lobe epilepsy. Annals of Neurology 57: 188–196

Concha L, Gross DW, Beaulieu C (2005b) Diffusion tensor tractography of the limbic system. AJNR: American Journal of Neuroradiology 26: 2267–2274

Conturo TE, Lori NF, Cull TS, Akbudak E, Snyder AZ, Shimony JS, McKinstry RC, Burton H, Raichle ME (1999) Tracking neuronal fiber pathways in the living human brain. Proceedings of the National Academy of Sciences of the United States of America 96: 10422–10427

Conturo TE, McKinstry RC, Akbudak E, Robinson BH (1996) Encoding of anisotropic diffusion with tetrahedral gradients: a general mathematical diffusion formalism and experimental results. Magnetic Resonance in Medicine 35: 399–412

Ding Z, Gore JC, Anderson AW (2003) Classification and quantification of neuronal fiber pathways using diffusion tensor MRI. Magnetic Resonance in Medicine 49: 716–721

Dougherty RF, Ben-Shachar M, Bammer R, Brewer AA, Wandell BA (2005) Functional organization of human occipital-callosal fiber tracts. Proceedings of the National Academy of Sciences of the United States of America 102: 7350–7355

Duncan JS, Bartlett P, Barker GJ (1996) Technique for measuring hippocampal T2 relaxation time. AJNR: American Journal of Neuroradiology 17:1805–1810

Einstein A (1926) Investigations on the Theory of the Brownian Movement. Methuen and Co.

Ellis CM, Simmons A, Jones DK, Bland J, Dawson JM, Horsfield MA, Williams SC, Leigh PN (1999) Diffusion tensor MRI assesses corticospinal tract damage in ALS. Neurology 53: 1051–1058

Field AS, Alexander AL, Wu YC, Hasan KM, Witwer B, Badie B (2004) Diffusion tensor eigenvector directional color imaging patterns in the evaluation of cerebral white matter tracts altered by tumor. Journal of Magnetic Resonance Imaging 20: 555–562

Filippi M, Rocca MA, Falini A, Caputo D, Ghezzi A, Colombo B, Scotti G, Comi G (2002) Correlations between structural CNS damage and functional MRI changes in primary progressive MS. Neuroimage 15: 537–546

Flynn SW, Lang DJ, Mackay AL, Goghari V, Vavasour IM, Whittall KP, Smith GN, Arango V, Mann JJ, Dwork AJ, Falkai P, Honer WG (2003) Abnormalities of myelination in schizophrenia detected in vivo with MRI, and post-mortem with analysis of oligodendrocyte proteins. Molecular Psychiatry 8: 811–820

Foong J, Symms MR, Barker GJ, Maier M, Woermann FG, Miller DH, Ron MA (2001) Neuropathological abnormalities in schizophrenia: evidence from magnetization transfer imaging. Brain 124 (Pt 5): 882–892

Frank LR (2002) Characterization of anisotropy in high angular resolution diffusion-weighted MRI. Magnetic Resonance in Medicine 47: 1083–1099

Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ (1993) Functional connectivity: the principal-component analysis of large (PET) data sets. Journal of Cerebral Blood Flow and Metabolism 13: 5–14

Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. Neuroimage 19: 1273–1302

Gareau PJ, Rutt BK, Karlik SJ, Mitchell JR (2000) Magnetization transfer and multicomponent T2 relaxation measurements with histopathologic correlation in an experimental model of MS. Journal of Magnetic Resonance Imaging 11: 586–595

Good CD, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS (2001) A voxel-based morphometric study of ageing in 465 normal adult human brains. Neuroimage 14: 21–36

Gudbjartsson H, Maier SE, Jolesz FA (1997) Double line scan diffusion imaging. Magnetic Resonance in Medicine 38: 101–109

Guye M, Parker GJ, Symms M, Boulby P, Wheeler-Kingshott CA, Salek-Haddadi A, Barker GJ, Duncan JS (2003) Combined functional MRI and tractography to demonstrate the connectivity of the human primary motor cortex in vivo. Neuroimage 19: 1349–1360

Hahn E (1950) Spin Echoes. Physics Review 80: 580–594

Hasan KM, Basser PJ, Parker DL, Alexander AL (2001a) Analytical computation of the eigenvalues and eigenvectors in DT-MRI. Journal of Magnetic Resonance 152: 41–47

Hasan KM, Parker DL, Alexander AL (2001b) Comparison of gradient encoding schemes for diffusion-tensor MRI. Journal of Magnetic Resonance Imaging 13: 769–780

Haselgrove JC, Moore JR (1996) Correction for distortion of echo-planar images used to calculate the apparent diffusion coefficient. Magnetic Resonance in Medicine 36: 960–964

Heiervang E, Behrens TE, Mackay CE, Robson MD, Johansen-Berg H (2006) Between session reproducibility and between subject variability of diffusion MR and tractography measures. Neuroimage 33: 867–877

Henkelman RM, Huang X, Xiang QS, Stanisz GJ, Swanson SD, Bronskill MJ (1993) Quantitative interpretation of magnetization transfer. Magnetic Resonance in Medicine 29: 759–766

Henkelman RM, Stanisz GJ, Graham SJ (2001) Magnetization transfer in MRI: a review. NMR in Biomedicine 14: 57–64

Henry RG, Berman JI, Nagarajan SS, Mukherjee P, Berger MS (2004) Subcortical pathways serving cortical language sites: initial experience with diffusion tensor imaging fiber tracking combined with intraoperative language mapping. Neuroimage 21: 616–622

Hickman SJ, Toosy AT, Jones SJ, Altmann DR, Miszkiel KA, MacManus DG, Barker GJ, Plant GT, Thompson AJ, Miller DH (2004) Serial magnetization transfer imaging in acute optic neuritis. Brain 127: 692–700

Holodny AI, Schwartz TH, Ollenschleger M, Liu WC, Schulder M (2001) Tumor involvement of the corticospinal tract: diffusion magnetic resonance tractography with intraoperative correlation. Journal of Neurosurgery 95: 1082

Horsfield MA, Barker GJ, Barkhof F, Miller DH, Thompson AJ, Filippi M (2003) Guidelines for using quantitative magnetization transfer magnetic resonance imaging for monitoring treatment of multiple sclerosis. Journal of Magnetic Resonance Imaging 17: 389–397

Horwitz B, Duara R, Rapoport SI (1984) Intercorrelations of glucose metabolic rates between brain regions: application to healthy males in a state of reduced sensory input. Journal of Cerebral Blood Flow and Metabolism 4: 484–499

Horwitz B, Tagamets MA, McIntosh AR (1999) Neural modeling, functional brain imaging, and cognition. Trends in Cognitive Sciences 3: 91–98

Jacob S, Finsterbusch J, Weishaupt JH, Khorram-Sefat D, Frahm J, Ehrenreich H (2003) Diffusion tensor imaging for long-term follow-up of corticospinal tract degeneration in amyotrophic lateral sclerosis. Neuroradiology 45: 598–600

Jellison BJ, Field AS, Medow J, Lazar M, Salamat MS, Alexander AL (2004) Diffusion tensor imaging of cerebral white matter: a pictorial review of physics, fiber tract anatomy, and tumor imaging patterns. AJNR: American Journal of Neuroradiology 25: 356–369

Jernigan TL, Archibald SL, Berhow MT, Sowell ER, Foster DS, Hesselink JR (1991) Cerebral structure on MRI, Part I: Localization of age-related changes. Biological Psychiatry 29: 55–67

Jezzard P, Balaban RS (1995) Correction for geometric distortion in echo planar images from B0 field variations. Magnetic Resonance in Medicine 34: 65–73

Johansen-Berg H, Behrens TE, Robson MD, Drobnjak I, Rushworth MF, Brady JM, Smith SM, Higham DJ, Matthews PM (2004) Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. Proceedings of the National Academy of Sciences of the United States of America 101: 13335–13340

Johansen-Berg H, Behrens TE, Sillery E, Ciccarelli O, Thompson AJ, Smith SM, Matthews PM (2005) Functional-anatomical validation and individual variation of diffusion tractography-based segmentation of the human thalamus. Cerebral Cortex 15: 31–39

Jones CK, Xiang QS, Whittall KP, MacKay AL (2004) Linear combination of multiecho data: short T2 component selection. Magnetic Resonance in Medicine 51: 495–502

Jones DK (2003) Determining and Visualizing Uncertainty in Estimates of Fiber Orientation From Diffusion Tensor MRI. Magnetic Resonance in Medicine 49: 7–12

Jones DK (2004) The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: a Monte Carlo study. Magnetic Resonance in Medicine 51: 807–815

Jones DK, Catani M, Pierpaoli C, Reeves SJ, Shergill SS, O'Sullivan M, Golesworthy P, McGuire P, Horsfield MA, Simmons A, Williams SC, Howard RJ (2006) Age effects on diffusion tensor magnetic resonance imaging tractography measures of frontal cortex connections in schizophrenia. Human Brain Mapping 27: 230–238

Jones DK, Catani M, Pierpaoli C, Reeves SJ, Shergill SS, O'Sullivan M, Maguire P, Horsfield MA, Simmons A, Williams SC, Howard RJ (2005a) A diffusion tensor magnetic resonance imaging study of frontal cortex connections in very-late-onset schizophrenia-like psychosis. American Journal of Geriatric Psychiatry 13: 1092–1099

Jones DK, Griffin LD, Alexander DC, Catani M, Horsfield MA, Howard R, Williams SC (2002a) Spatial normalization and averaging of diffusion tensor MRI data sets. Neuroimage 17: 592–617

Jones DK, Horsfield MA, Simmons A (1999) Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. Magnetic Resonance in Medicine 42: 515–525

Jones DK, Symms MR, Cercignani M, Howard RJ (2005b) The effect of filter size on VBM analyses of DT-MRI data. Neuroimage 26: 546–554

Jones DK, Travis AR, Eden G, Pierpaoli C, Basser PJ (2005c) PASTA: pointwise assessment of streamline tractography attributes. Magnetic Resonance in Medicine 53: 1462–1467

Jones DK, Williams SC, Gasston D, Horsfield MA, Simmons A, Howard R (2002b) Isotropic resolution diffusion tensor imaging with whole brain acquisition in a clinically acceptable time. Human Brain Mapping 15: 216–230

Kabani NJ, Sled JG, Chertkow H (2002a) Magnetization transfer ratio in mild cognitive impairment and dementia of Alzheimer's type. Neuroimage 15: 604–610

Kabani NJ, Sled JG, Shuper A, Chertkow H (2002b) Regional magnetization transfer ratio changes in mild cognitive impairment. Magnetic Resonance in Medicine 47: 143–148

Kalus P, Slotboom J, Gallinat J, Wiest R, Ozdoba C, Federspiel A, Strik WK, Buri C, Schroth G, Kiefer C (2005) The amygdala in schizophrenia: a trimodal magnetic resonance imaging study. Neuroscience Letters 375: 151–156

Kanaan RA, Shergill SS, Barker GJ, Catani M, Ng VW, Howard R, McGuire PK, Jones DK (2006) Tract-specific anisotropy measurements in diffusion tensor imaging. Psychiatry Research 146: 73–82

Kiebel SJ, Poline JB, Friston KJ, Holmes AP, Worsley KJ (1999) Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. Neuroimage 10: 756–766

Kiefer C, Slotboom J, Buri C, Gralla J, Remonda L, Dierks T, Strik WK, Schroth G, Kalus P (2004) Differentiating hippocampal subregions by means of quantitative magnetization transfer and relaxometry: preliminary results. Neuroimage 23: 1093–1099

Kinoshita M, Yamada K, Hashimoto N, Kato A, Izumoto S, Baba T, Maruno M, Nishimura T, Yoshimine T (2005) Fiber-tracking does not accurately estimate size of fiber bundle in pathological condition: initial neurosurgical experience using neuronavigation and subcortical white matter stimulation. Neuroimage 25: 424–429

Kubicki M, Park H, Westin CF, Nestor PG, Mulkern RV, Maier SE, Niznikiewicz M, Connor EE, Levitt JJ, Frumin M, Kikinis R, Jolesz FA, McCarley RW, Shenton ME (2005) DTI and MTR abnormalities in schizophrenia: analysis of white matter integrity.

Kumar A, Gupta RC, Albert Thomas M, Alger J, Wyckoff N, Hwang S (2004) Biophysical changes in normal-appearing white matter and subcortical nuclei in late-life major depression detected using magnetization transfer. Psychiatry Research 130: 131–140

Laule C, Vavasour IM, Moore GR, Oger J, Li DK, Paty DW, MacKay AL (2004) Water content and myelin water fraction in multiple sclerosis. A T2 relaxation study. Journal of Neurology 251: 284–293

Lazar M, Alexander AL (2002) White matter tractography using random vector (RAVE) perturbation. Proceedings of 10th Annual Meeting of the ISMRM, Honolulu, Hawaii: 539

Lazar M, Alexander AL (2005) Bootstrap white matter tractography (BOOT-TRAC). Neuroimage 24: 524–532

Lazar M, Alexander AL, Thottakara P, Badie B, Field AS (2006) White matter reorganization after surgical resection of brain tumors and vascular malformations. AJNR: American Journal of Neuroradiology: in press

Lazar M, Weinstein DM, Tsuruda JS, Hasan KM, Arfanakis K, Meyerand ME, Badie B, Rowley HA, Haughton V, Field A, Alexander AL (2003) White matter tractography using diffusion tensor deflection. Human Brain Mapping 18: 306–321

Le Bihan D (1990) Diffusion/perfusion MR imaging of the brain: from structure to function. Radiology 177: 328–329

Lim KO, Hedehus M, Moseley M, de Crespigny A, Sullivan EV, Pfefferbaum A (1999) Compromised white matter tract integrity in schizophrenia inferred from diffusion tensor imaging. Archives of General Psychiatry 56: 367–374

Lin X, Tench CR, Morgan PS, Niepel G, Constantinescu CS (2005) 'Importance sampling' in MS: use of diffusion tensor tractography to quantify pathology related to specific impairment. Journal of the Neurological Sciences 237: 13–19

Liu C, Bammer R, Acar B, Moseley ME (2004) Characterizing non-Gaussian diffusion by using generalized diffusion tensors. Magnetic Resonance in Medicine 51: 924–937

MacKay A, Laule C, Vavasour I, Bjarnason T, Kolind S, Madler B (2006) Insights into brain microstructure from the T2 distribution. Magnetic resonance imaging. 24: 515–525

MacKay A, Whittall K, Adler J, Li D, Paty D, Graeb D (1994) In vivo visualization of myelin water in brain by magnetic resonance. Magnetic Resonance in Medicine 31: 673–677

Madden DJ, Whiting WL, Huettel SA, White LE, MacFall JR, Provenzale JM (2004) Diffusion tensor imaging of adult age differences in cerebral white matter: relation to response time. Neuroimage 21: 1174–1181

Maier SE, Vajapeyam S, Mamata H, Westin CF, Jolesz FA, Mulkern RV (2004) Biexponential diffusion tensor analysis of human brain diffusion data. Magnetic Resonance in Medicine 51: 321–330

Makris N, Worth AJ, Sorensen AG, Papadimitriou GM, Wu O, Reese TG, Wedeen VJ, Davis TL, Stakes JW, Caviness VS, Kaplan E, Rosen BR, Pandya DN, Kennedy DN (1997) Morphometry of in vivo human white matter association pathways with diffusion-weighted magnetic resonance imaging. Annals of Neurology 42: 951–962

Mansfield P (1984) Real-time echo-planar imaging by NMR. British Medical Bulletin 40: 187–190

Martin, Kikinis R, Zuerrer M, Boesch C, Briner J, Kewity G, Kaelin P (1988) Developmental stages of human brain: an MR study. Journal of Computer Assisted Tomography 12: 917–922

McIntosh AR (2000) Towards a network theory of cognition. Neural Networks 13: 861–870

Menon RS, Allen PS (1991) Application of continuous relaxation time distributions to the fitting of data from model systems and excised tissue. Magnetic Resonance in Medicine 20: 214–227

Miot-Noirault E, Barantin L, Akoka S, Le Pape A (1997) T2 relaxation time as a marker of brain myelination: experimental MR study in two neonatal animal models. Journal of Neuroscience Methods 72: 5–14

Miot E, Hoffschir D, Poncy JL, Masse R, Le Pape A, Akoka S (1995) Magnetic resonance imaging in vivo monitoring of T2 relaxation time: quantitative assessment of primate brain maturation. Journal of Medical Primatology 24: 87–93

Mori S, Crain BJ, Chacko VP, van Zijl PCM (1999) Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. Annals of Neurology 45: 265–269

Moseley ME, Cohen Y, Kucharczyk J, Mintorovitch J, Asgari HS, Wendland MF, Tsuruda J, Norman D (1990) Diffusion-weighted MR imaging of anisotropic water diffusion in cat central nervous system. Radiology 176: 439–445

Mulkern RV, Gudbjartsson H, Westin CF, Zengingonul HP, Gartner W, Guttmann CR, Robertson RL, Kyriakos W, Schwartz R, Holtzman D, Jolesz FA, Maier SE (1999) Multi-component apparent diffusion coefficients in human brain. NMR in Biomedicine 12: 51–62

Neil J, Miller J, Mukherjee P, Huppi PS (2002) Diffusion tensor imaging of normal and injured developing human brain - a technical review. NMR in Biomedicine 15: 543–552

Niendorf T, Dijkhuizen RM, Norris DG, van Lookeren Campagne M, Nicolay K (1996) Biexponential diffusion attenuation in various states of brain tissue: implications for diffusion-weighted imaging. Magnetic Resonance in Medicine 36: 847–857

Oh J, Han ET, Pelletier D, Nelson SJ (2006) Measurement of in vivo multi-component T2 relaxation times for brain tissue using multi-slice T2 prep at 1.5 and 3 T. Magnetic Resonance Imaging 24: 33–43

Olesen PJ, Nagy Z, Westerberg H, Klingberg T (2003) Combined analysis of DTI and fMRI data reveals a joint maturation of white and grey matter in a fronto-parietal network. Cognitive Brain Research 18: 48–57

Ozarslan E, Mareci TH (2003) Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. Magnetic Resonance in Medicine 50: 955–965

Pajevic S, Pierpaoli C (1999) Color schemes to represent the orientation of anisotropic tissues from diffusion tensor data: application to white matter fiber tract mapping in the human brain. Magnetic Resonance in Medicine 42: 526–540

Papadakis NG, Martin KM, Mustafa MH, Wilkinson ID, Griffiths PD, Huang CL, Woodruff PW (2002) Study of the effect of CSF suppression on white matter diffusion anisotropy mapping of healthy human brain. Magnetic Resonance in Medicine 48: 394–398

Papadakis NG, Xing D, Huang CL, Hall LD, Carpenter TA (1999) A comparative study of acquisition schemes for diffusion tensor imaging using MRI. Journal of Magnetic Resonance 137: 67–82

Park H-J, Kubicki M, Shenton ME, Guimond A, McCarley RW, Maier SE, Kikinis R, Jolesz FA, Westin CF (2003) Spatial normalization of diffusion tensor MRI using multiple channels. Neuroimage 20: 1995–2009

Parker GJ, Haroon HA, Wheeler-Kingshott CA (2003) A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurements. Journal of Magnetic Resonance Imaging 18: 242–254

Parker GJ, Stephan KE, Barker GJ, Rowe JB, MacManus DG, Wheeler-Kingshott CA, Ciccarelli O, Passingham RE, Spinks RL, Lemon RN, Turner R (2002) Initial demonstration of in vivo tracing of axonal projections in the macaque brain and comparison with the human brain using diffusion tensor imaging and fast marching tractography. Neuroimage 15: 797–809

Paus T, Collins DL, Evans AC, Leonard G, Pike B, Zijdenbos A (2001) Maturation of white matter in the human brain: a review of magnetic resonance studies. Brain Research Bulletin 54: 255–266

Penny WD, Stephan KE, Mechelli A, Friston KJ (2004) Modelling functional integration: a comparison of structural equation and dynamic causal models. Neuroimage 23 Suppl 1: S264–274

Pfefferbaum A, Adalsteinsson E, Sullivan EV (2005) Frontal circuitry degradation marks healthy adult aging: Evidence from diffusion tensor imaging. Neuroimage 26: 891–899

Pierpaoli C, Basser PJ (1996) Toward a quantitative assessment of diffusion anisotropy. Magnetic Resonance in Medicine 36: 893–906

Pipe JG, Farthing VG, Forbes KP (2002) Multishot diffusion-weighted FSE using PROPELLER MRI.[erratum appears in Magn Reson Med 2002 Mar;47(3):621]. Magnetic Resonance in Medicine 47: 42–52

Pomara N, Crandall DT, Choi SJ, Johnson G, Lim KO (2001) White matter abnormalities in HIV-1 infection: a diffusion tensor imaging study. Psychiatry Research 106: 15–24

Poon CS, Henkelman RM (1992) Practical T2 quantitation for clinical applications. Journal of Magnetic Resonance Imaging 2: 541–553

Poon CS, Henkelman RM (1995) Robust refocusing pulses of limited power. Journal of Magnetic Resonance 116: 161–180

Prayer D, Prayer L (2003) Diffusion-weighted magnetic resonance imaging of cerebral white matter development. European Journal of Radiology 45: 235–243

Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P (1999) SENSE: sensitivity encoding for fast MRI. Magnetic Resonance in Medicine 42: 952–962

Rademacher J, Engelbrecht V, Burgel U, Freund H, Zilles K (1999) Measuring in vivo myelination of human white matter fiber tracts with magnetization transfer MR. Neuroimage 9: 393–406

Ragin AB, Storey P, Cohen BA, Epstein LG, Edelman RR (2004) Whole brain diffusion tensor imaging in HIV-associated cognitive impairment. AJNR: American Journal of Neuroradiology 25: 195–200

Ramnani N, Behrens TE, Penny W, Matthews PM (2004) New approaches for exploring anatomical and functional connectivity in the human brain. Biological Psychiatry 56: 613–619

Reese TG, Heid O, Weisskoff RM, Wedeen VJ (2003) Reduction of eddy-current-induced distortion in diffusion MRI using a twice-refocused spin echo. Magnetic Resonance in Medicine 49: 177–182

Rocca MA, Agosta F, Martinelli V, Falini A, Comi G, Filippi M (2005) The level of spinal cord involvement influences the pattern of movement-associated cortical recruitment in patients with isolated myelitis. Neuroimage

Rohde GK, Barnett AS, Basser PJ, Marenco S, Pierpaoli C (2004) Comprehensive approach for correction of motion and distortion in diffusion-weighted MRI. Magnetic Resonance in Medicine 51: 103–114

Rohde GK, Barnett AS, Basser PJ, Pierpaoli C (2005) Estimating intensity variance due to noise in registered images: applications to diffusion tensor MRI. Neuroimage 26: 673–684

Ronen I, Ugurbil K, Kim DS (2005) How does DWI correlate with white matter structures? Magnetic Resonance in Medicine 54: 317–323

Ropele S, Filippi M, Valsasina P, Korteweg T, Barkhof F, Tofts PS, Samson R, Miller DH, Fazekas F (2005) Assessment and correction of B1-induced errors in magnetization transfer ratio measurements. Magnetic Resonance in Medicine 53: 134–140

Rovaris M, Iannucci G, Falautano M, Possa F, Martinelli V, Comi G, Filippi M (2002) Cognitive dysfunction in patients with mildly disabling relapsing-remitting multiple sclerosis: an exploratory study with diffusion tensor MR imaging. Journal of the Neurological Sciences 195: 103–109

Rushworth MF, Behrens TE, Johansen-Berg H (2005) Connection Patterns Distinguish 3 Regions of Human Parietal Cortex. Cerebral Cortex: epub ahead of print

Salat DH, Tuch DS, Greve DN, van der Kouwe AJ, Hevelone ND, Zaleta AK, Rosen BR, Fischl B, Corkin S, Rosas HD, Dale AM (2005) Age-related alterations in white matter microstructure measured by diffusion tensor imaging. Neurobiology of Aging 26: 1215–1227

Salmond CH, Menon DK, Chatfield DA, Williams GB, Pena A, Sahakian BJ, Pickard JD (2006) Diffusion tensor imaging in chronic head injury survivors: correlations with learning and memory indices. Neuroimage 29: 117–124

Salonen O, Autti T, Raininko R, Ylikoski A, Erkinjuntti T (1997) MRI of the brain in neurologically healthy middle-aged and elderly individuals. Neuroradiology 39: 537–545

Santyr GE (1993) Magnetization transfer effects in multislice MR imaging. Magnetic Resonance Imaging 11: 521–532

Shimony JS, McKinstry RC, Akbudak E, Aronovitz JA, Snyder AZ, Lori NF, Cull TS, Conturo TE (1999) Quantitative diffusion-tensor anisotropy brain MR imaging: normative human data and anatomic analysis. Radiology 212: 770–784

Shrager RI, Basser PJ (1998) Anisotropically weighted MRI. Magnetic Resonance in Medicine 40: 160–165

Sie LT, van der Knaap MS, van Wezel-Meijler G, Valk J (1997) MRI assessment of myelination of motor and sensory pathways in the brain of preterm and term-born infants. Neuropediatrics 28: 97–105

Skare S, Andersson JL (2001) On the effects of gating in diffusion imaging of the brain using single shot EPI. Magnetic Resonance Imaging 19: 1125–1128

Skare S, Li T-Q, Nordell B, Ingvar M (2000) Noise considerations in the determination of diffusion tensor anisotropy. Magnetic Resonance Imaging 18: 659–669

Sled JG, Pike B (2000) Correction for B1 and B0 variations in quantitative T2 measurements using MRI. Magnetic Resonance in Medicine 43: 589–593

Sled JG, Pike GB (2001) Quantitative imaging of magnetization transfer exchange and relaxation properties in vivo using MRI. Magnetic Resonance in Medicine 46: 923–931

Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay C, Watkins KE, Ciccarelli O, Cader MZ, M. MP, E.J. B (2006) Tract-Based Spatial Statistics: Voxelwise Analysis of Multi-Subject Diffusion Data. Neuroimage: under review

Snook L, Paulson LA, Roy D, Phillips L, Beaulieu C (2005) Diffusion tensor imaging of neurodevelopment in children and young adults. Neuroimage 26: 1164–1173

Song SK, Sun SW, Ramsbottom MJ, Chang C, Russell J, Cross AH (2002) Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water. Neuroimage 17: 1429–1436

Song SK, Yoshino J, Le TQ, Lin SJ, Sun SW, Cross AH, Armstrong RC (2005) Demyelination increases radial diffusivity in corpus callosum of mouse brain. Neuroimage 26: 132–140

Stanisz GJ, Henkelman RM (1998) Diffusional anisotropy of T2 components in bovine optic nerve. Magnetic Resonance in Medicine 40: 405–410

Stanisz GJ, Kecojevic A, Bronskill MJ, Henkelman RM (1999) Characterizing white matter with magnetization transfer and T. Magnetic Resonance in Medicine 42: 1128–1136

Stanisz GJ, Webb S, Munro CA, Pun T, Midha R (2004) MR properties of excised neural tissue following experimentally induced inflammation. Magnetic Resonance in Medicine 51: 473–479

Steen RG, Ogg RJ, Reddick WE, Kingsley PB (1997) Age-related changes in the pediatric brain: quantitative MR evidence of maturational changes during adolescence. AJNR: American Journal of Neuroradiology 18: 819–828

Stejskal E, Tanner J (1965) Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. Journal of Chemical Physics 42: 288–292

Stephan KE (2004) On the role of general system theory for functional neuroimaging. Journal of Anatomy 205: 443–470

Stephan KE, Harrison LM, Penny WD, Friston KJ (2004) Biophysical models of fMRI responses. Current Opinion in Neurobiology 14: 629–635

Stephan KE, Penny WD, Marshall JC, Fink GR, Friston KJ (2005) Investigating the functional role of callosal connections with dynamic causal models. Annals of the New York Academy of Science 1064: 16–36

Stieltjes B, Kaufmann WE, van Zijl PC, Fredericksen K, Pearlson GD, Solaiyappan M, Mori S (2001) Diffusion tensor imaging and axonal tracking in the human brainstem. Neuroimage 14: 723–735

Szeszko PR, Ardekani BA, Ashtari M, Malhotra AK, Robinson DG, Bilder RM, Lim KO (2005) White matter abnormalities in obsessive-compulsive disorder: a diffusion tensor imaging study. Archives of General Psychiatry 62: 782–790

Thottakara P, Lazar M, Johnson SC, Alexander AL (2006) Application of Brodmann's area templates for ROI selection in white matter tractography studies. Neuroimage 29: 868–878

Toosy AT, Ciccarelli O, Parker GJ, Wheeler-Kingshott CA, Miller DH, Thompson AJ (2004) Characterizing function-structure relationships in the human visual system with functional MRI and diffusion tensor imaging. Neuroimage 21: 1452–1463

Toosy AT, Werring DJ, Orrell RW, Howard RS, King MD, Barker GJ, Miller DH, Thompson AJ (2003) Diffusion tensor imaging detects corticospinal tract involvement at multiple levels in amyotrophic lateral sclerosis. Journal of Neurology, Neurosurgery & Psychiatry 74: 1250–1257

Townsend TN, Bernasconi N, Pike GB, Bernasconi A (2004) Quantitative analysis of temporal lobe white matter T2 relaxation time in temporal lobe epilepsy. Neuroimage 23: 318–324

Tozer D, Ramani A, Barker GJ, Davies GR, Miller DH, Tofts PS (2003) Quantitative magnetization transfer mapping of bound protons in multiple sclerosis. Magnetic Resonance in Medicine 50: 83–91

Tozer DJ, Davies GR, Altmann DR, Miller DH, Tofts PS (2005) Correlation of apparent myelin measures obtained in multiple sclerosis patients and controls from magnetization transfer and multicompartmental T2 analysis. Magnetic Resonance in Medicine 53: 1415–1422

Tuch DS (2004) Q-ball imaging. Magnetic Resonance in Medicine 52: 1358–1372

Tuch DS, Salat DH, Wisco JJ, Zaleta AK, Hevelone ND, Rosas HD (2005) Choice reaction time performance correlates with diffusion anisotropy in white matter pathways supporting visuospatial attention. Proceedings of the National Academy of Sciences of the United States of America 102: 12212–12217

Tuch DS, Wedeen VJ, Dale AM, George JS, Belliveau JW (2001) Conductivity tensor mapping of the human brain using diffusion tensor MRI. Proceedings of the National Academy of Sciences of the United States of America 98: 11697–11701

Turner R, Le Bihan D, Maier J, Vavrek R, Hedges LK, Pekar J (1990) Echo-planar imaging of intravoxel incoherent motion. Radiology 177: 407–414

Ulug AM, van Zijl PC (1999) Orientation-independent diffusion imaging without tensor diagonalization: anisotropy definitions based on physical attributes of the diffusion ellipsoid. Journal of Magnetic Resonance Imaging 9: 804–813

Vaithianathar L, Tench CR, Morgan PS, Wilson M, Blumhardt LD (2002) T1 relaxation time mapping of white matter tracts in multiple sclerosis defined by diffusion tensor imaging. Journal of Neurology 249: 1272–1278

van Buchem MA, Steens SC, Vrooman HA, Zwinderman AH, McGowan JC, Rassek M, Engelbrecht V (2001) Global estimation of myelination in the developing brain on the basis of magnetization transfer imaging: a preliminary study. AJNR: American Journal of Neuroradiology 22: 762–766

Vavasour IM, Whittall KP, MacKay AL, Li DK, Vorobeychik G, Paty DW (1998) A comparison between magnetization transfer ratios and myelin water percentages in normals and multiple sclerosis patients. Magnetic Resonance in Medicine 40: 763–768

Vidarsson L, Conolly SM, Lim KO, Gold GE, Pauly JM (2005) Echo time optimization for linear combination myelin imaging. Magnetic Resonance in Medicine 53: 398–407

Wakana S, Jiang H, Nagae-Poetscher LM, van Zijl PC, Mori S (2004) Fiber tract-based atlas of human white matter anatomy. Radiology 230: 77–87

Webb S, Munro CA, Midha R, Stanisz GJ (2003) Is multicomponent T2 a good measure of myelin content in peripheral nerve? Magnetic Resonance in Medicine 49: 638–645

Wedeen VJ, Hagmann P, Tseng WY, Reese TG, Weisskoff RM (2005) Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. Magnetic Resonance in Medicine 54: 1377–1386

Werring DJ, Clark CA, Barker GJ, Miller DH, Parker GJ, Brammer MJ, Bullmore ET, Giampietro VP, Thompson AJ (1998) The structural and functional mechanisms of motor recovery: complementary use of diffusion tensor and functional magnetic resonance imaging in a traumatic injury of the internal capsule. Journal of Neurology, Neurosurgery & Psychiatry 65: 863–869

Westin CF, Maier SE, Mamata H, Nabavi A, Jolesz FA, Kikinis R (2002) Processing and visualization for diffusion tensor MRI. Medical Image Analysis 6: 93–108

Whittall KP, MacKay AL, Graeb DA, Nugent RA, Li DK, Paty DW (1997) In vivo measurement of T2 distributions and water contents in normal human brain. Magnetic Resonance in Medicine 37: 34–43

Whittall KP, MacKay AL, Li DK, Vavasour IM, Jones CK, Paty DW (2002) Normal-appearing white matter in multiple sclerosis has heterogeneous, diffusely prolonged T(2). Magnetic Resonance in Medicine 47: 403–408

Whittall KP, MacKay AL, Li DKB (1999) Are mono-exponential fits to a few echoes sufficient to determine T2 relaxation for in-vivo human brain? Magnetic Resonance in Medicine 41: 1255–1257

Witwer BP, Moftakhar R, Hasan KM, Deshmukh P, Haughton V, Field A, Arfanakis K, Noyes J, Moritz CH, Meyerand ME, Rowley HA, Alexander AL,

Badie B (2002) Diffusion-tensor imaging of white matter tracts in patients with cerebral neoplasm. Journal of Neurosurgery 97: 568–575

Wu YC, Alexander AL (2005) Hybrid diffusion imaging for complex diffusion characterization. Proceedings of the 13th ISMRM Scientific Meeting: 578

Wu YC, Field AS, Chung MK, Badie B, Alexander AL (2004) Quantitative analysis of diffusion tensor orientation: theoretical framework. Magnetic Resonance in Medicine 52: 1146–1155

Xu D, Mori S, Shen D, van Zijl PC, Davatzikos C (2003) Spatial normalization of diffusion tensor fields. Magnetic Resonance in Medicine 50: 175–182

Yarnykh VL (2002) Pulsed Z-spectroscopic imaging of cross-relaxation parameters in tissues for human MRI: theory and clinical applications. Magnetic Resonance in Medicine 47: 929–939

Yarnykh VL, Yuan C (2004) Cross-relaxation imaging reveals detailed anatomy of white matter fiber tracts in the human brain. Neuroimage 23: 409–424

Zhang Y, Wehrli FW (2004) Reference-scan-free method for automated correction of Nyquist ghost artifacts in echoplanar brain images. Magnetic Resonance in Medicine 51: 621–624

Part III

Analysis and Measures of Brain Connectivity

Simulation Frameworks for Large-Scale Brain Systems

Barry Horwitz and Fatima T Husain

Brain Imaging & Modeling Section, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD 20892

In this paper we review the relatively recent effort on the part of cognitive neuroscientists to use computational neural modeling in conjunction with functional brain imaging, especially the hemodynamic-based methods such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET). The reason why such a review is in a book on brain connectivity is that many of these efforts involve neural models that consist of multiple interacting neuronal populations, and thus issues associated with connectivity are often implicitly or explicitly addressed.

1 Neural Modeling – an Introduction

The term neural modeling refers to a variety of different computational schemes (Arbib 2003), a spectrum if you will, ranging from those using perceptrons and backpropagation (McClelland & Rumelhart 1986) that often lack biological plausibility, to neural neworks that incorporate elements with biologically realistic properties (Dayan & Abbott 2001). Although, as we shall see, models at various points along this spectrum are now being utilized in conjunction with neuroimaging data, this review will emphasize the use of biologically realistic neural models.

Until recently, the focus of most mammalian neural modeling work centered on the behavior of single neurons, or small populations of neurons, usually located in a single brain area (Arbib 2003; Dayan & Abbott 2001; Rolls & Treves 1998), although there were some important exceptions, such as Tononi et al. (1992). The reason for such a focus was clear: most neural data suitable for modeling were acquired from single unit electrophysiological recordings. Beginning in the early 1990s, investigators started employing various types of computational network modeling methods that were directed at functional brain imaging data (e.g., Arbib et al. 1995; Friston 1994; Friston et al. 1991; Horwitz 1990; McIntosh & Gonzalez-Lima 1991; McIntosh et al. 1994; Tagamets & Horwitz 1998). Hemodynamic-based functional brain

imaging has the ability to provide information about brain activity during the performance of cognitive tasks with a spatial resolution ranging from several millimeters to 1–2 centimeters from essentially all brain regions simultaneously. Moreover, because these methods are relatively non-invasive and can be performed in normal subjects, they enable brain researchers to investigate the brain basis of human cognition (for a review, see Frackowiak et al. 2004). However, the temporal resolution of these types of data are much inferior to the millisecond scale of neural dynamics (for PET, the temporal resolution is about 30–60 sec; for fMRI, it is several seconds). Other techniques, such as electroencephalography (EEG) and magnetoencephalograpy (MEG) can also be used, and these methods do provide the requisite temporal information, but spatial localization is less well defined than is the case for fMRI and PET (see Horwitz et al. (2000) and Horwitz & Poeppel (2002) for brief discussions of the various neuroimaging methods and for the difficulties in combining them). Nonetheless, EEG/MEG has also elicited a number of computational neural modeling efforts (e.g., David & Friston 2003; Jirsa & Haken 1997; May et al. 1999; Nunez 1981; Robinson et al. 2005).

The central role that functional neuroimaging now plays in human cognitive neuroscience cannot be emphasized enough. Although there are numerous tasks that can be similarly performed in humans and nonhumans (especially nonhuman primates), there are many cognitive functions that are difficult, if not impossible, to study in nonhumans, especially those related to language, to some aspects of social cognition, and to high level executive function. Until the advent of functional neuroimaging the only ways to investigate the neural basis of human cognition were: (1) analysis of the behavioral consequences of brain lesions (e.g., strokes); (2) electrophysiological studies in neurosurgical patients; (3) examination of behavior following pharmacologic intervention or in relation to genetic analysis; and, (4) extrapolation from nonhuman neurophysiological and other neuroscientific approaches. The hemodynamic functional neuroimaging methods (fMRI and PET) allowed numerous investigators to view the changes in brain activity between tasks and/or groups of subjects (e.g., normal volunteers and subjects with neurological or psychiatric disorders) with a spatial scale of a few millimeters, and, most importantly, to view these changes in most of the brain at the same time. The interpretation of these patterns of activity, involving the interaction of multiple and distributed neuronal populations, generated the need for computational modeling.

Modeling serves a number of purposes in this regard, including a way to keep track of the complex interactions between the various neural populations and a way to relate these patterns to neural mechanisms (e.g., Horwitz & Glabus 2005) or to hypothesized cognitive mechanisms (e.g., Anderson et al. 2003). Moreover, neural modeling is necessary, we would argue, for another reason. As indicated in the last paragraph, there are numerous and diverse sources of neuroscientific data that relate to the neural basis of cognitive function. All these types of data have different spatial, temporal and featural properties that make them hard to relate to one another. Furthermore, each

type of data has its own interpretational limitations. The net effect is that no one kind of neuroscientific data can be thought of as being a "gold standard". That is, all these different types of data (lesions, electrophysiological, functional neuroimaging, etc.) are providing us with some information about the neural basis of a cognitive function, but there is no easy and straightforward way to put all these types of information together. We have argued for a long time now (e.g., Horwitz et al. 1999) that computational neural modeling provides a method by which all relevant neuroscientific data pertaining to a cognitive task can be accounted for in terms of the dynamic interactions of multiple neuronal populations. We will illustrate this later in this paper.

Conceptually, there are two different ways to employ computational modeling, although both can be used in conjunction with one another. In simulation mode, a model is constructed (i.e., a set of model parameters is defined and values for each parameter are assigned) and data are generated from each of the elements of the model. These data are then compared with appropriate experimental data and the model is considered successful if there is close agreement between the experimental and simulated data. In data-fitting mode, some computational procedure is used to vary the model parameters until there is agreement between experimental data and data generated by the model. Examples of data-fitting modeling are (1) the use of Structural Equation Modeling (SEM) with PET or fMRI data (e.g., Buechel et al. 1999; McIntosh et al. 1994) and (2) the use of Dynamic Causal Model with fMRI data (Friston et al. 2003) (see also Stephan and Friston, this volume). Several papers in this volume focus on data-fitting modeling. Although the distinction between simulation and data-fitting is fuzzy, it is often the case that the models used in data-fitting are less detailed and less specific about the model parameters than are the models used in simulation. Importantly, the parameters used in simulation models often have their values based on a different set of data than on the data to which the model is being applied.

Simulation modeling can also be used directly at the level of PET/fMRI data (e.g., using SEM to simulate new results; see Horwitz 1990; Kronhaus & Willshaw in press). However, its main use has been to relate neurobiological data to fMRI or PET results, or conversely, to relate the performance of cognitive models to such data. The focus of this paper is on the former, but we shall also briefly review the latter.

Another way to think about the different types of modeling is that some are "bottom-up" and some are "top-down". In a bottom-up approach, the data one tries to explain are at one level, and the explanatory variables are at a "lower" level (lower can mean such things as more fundamental, more microscopic, more basic). As will be shown, this means, for example, trying to account for fMRI data in terms of the activity of neurons. Importantly, the goal of such modeling is to propose neural mechanisms that result in particular cognitive functions. In essence, one wants the cognitive function to appear as an emergent phenomenon. In a top-down approach, the hypothesized mechanisms are cognitive, and the goal is to locate the brain regions

that implement these cognitive mechanisms. The way this is done is to find the brain region(s) whose fMRI signals behave as proposed by the model.

In the next section, we will review some recent top-down approaches. In the following section, we will review some bottom-up studies. We will end with some concluding comments.

2 Top-down Methods

Cognitive models make assumptions about the types of functions that mediate a cognitive task. In many cases these models do not have any relationship to the functional neuroanatomy of the brain. For example, a well-known cognitive model of reading (Coltheart et al. 2001) includes, among others, modules for visual analysis and for grapheme-phoneme conversion, although no attempt was made to link activity in these modules to fMI or PET data, or even to link them to specific brain regions based on lesion studies. Recently, however, such efforts have been made by imposing additional assumptions relating each cognitive function to specified brain regions. An early attempt at combining a cognitive model with fMRI data can be found in the work of Just et al. (1999), who used a computational (production) model of sentence comprehension called 4CAPS to explain how fMRI activation levels varied as a function of sentence complexity in three brain areas (Broca, Wernicke and dorsolateral prefrontal cortex). In their computational model, Just et al. proposed that resource utilization in a given unit of time in each component of the system corresponds to the amount of activation observed with the neuroimaging measure in the corresponding component during that time interval. Good agreement between the experimental number of activated voxels in Broca and Wernicke's areas and the predictions of their model for three types of sentences of different complexity were found.

A recent study by Anderson et al. (2003) represents another example of this type of modeling. They examined symbol manipulation tasks using a model called ACT-R, which contains a number of buffers. Somewhat different from the assumption used by Just et al. (1999), Anderson et al. (2003) proposed that the fMRI response in a brain area represents the integrated duration of time that a buffer is active. They showed that calculated fMRI activity in one buffer of the model (the imaginal buffer, which tracks changes in problem representation) predicted the fMRI response of a left parietal region, activity in a second buffer (the retrieval buffer) predicted activity in a left prefrontal region, and activity in a third buffer of the model (the manual buffer) was related to fMRI activity in a motor region. In a second study (Anderson et al. 2005), they extended the model to a more complex task, the Tower of Hanoi, and were able to explain latency data in move generation and the fMRI responses in the three aforementioned regions.

One difference between the approaches of Just and Anderson is that Anderson and colleagues (Anderson et al. 2003, 2005) assume that each

module of their model corresponds to a given brain region. Just et al. (1999) also ascribe specific cognitive functions to different brain regions, although more than one cognitive function can occur in a region and, conversely, they also assume that a given cognitive specialization may occur in more than one area, albeit with different degrees of efficiency.

Another example of this type of modeling involves an investigation of the role of the anterior cingulate in cognitive control. Brown and Braver (2005) used a cognitive computational model that embodied the hypothesis that the response of the anterior cingulate to a given task condition is proportional to the perceived likelihood of an error in that condition. Simulations performed using the model with a modified stop-signal task resulted in a pattern of behavioral performance that fitted human data. Moreover, the pattern of cingulate activity in the model across task conditions (as indexed by the neural firing rate) was qualitatively similar to fMRI activity in the anterior cingulate obtained in an event-related experimental study. They also constructed a second model that viewed the anterior cingulate as detecting conflict between incompatible response properties. They found that they could also fit this model to human behavioral data, but that the pattern of simulated activity of the anterior cingulate across conditions did not match the pattern of the fMRI data. The Brown-Braver study provides an interesting example of how functional brain imaging data can be used in conjunction with cognitive modeling. Namely, different and competing cognitive models may equally well fit the performance data that they typically aim to explain. By comparing the ability of competing models to also match functional neuroimaging data, one is able to select one of the competing models over the other.

The examples we have presented illustrate some of the limitations to this "top-down" approach. One important limitation is that each study we presented employed a different relationship between the activity of a module and fMRI activity. For Just et al. (1999), the fMRI signal was indexed by the rate of resource utilization. This is probably similar to the measure used by Brown and Braver (2005) (i.e., the neural firing rate). For Anderson and colleagues (2003), the fMRI signal is proportional to the integrated duration of time that a buffer is active. Note that although each of these relationships between model activity and fMRI signal is plausible, and all are probably similar, there is no experimental way to verify the relation between a cognitive model component and a neural/hemodynamic variable. As we shall see in the next section, one advantage of a bottom-up approach is that it allows one to relate neural activity to the hemodynamic signal in a testable way. That is, assumptions about the relationship between neural activity and fMRI activity are also expressed in biological, not cognitive or functional, terms.

On the other hand, a big advantage, at least at the present time, for employing cognitive models in conjunction with functional neuroimaging data is that only relatively low-level cognitive functions can be readily addressed in neural terms, since non-human models of high-level human cognitive function (e.g., language) do not exist. The use of cognitive models with fMRI data

allows one to deal with very high-level cognitive phenomena (e.g., sentence processing, symbol manipulation), but gives little indication as to how such functions are implemented at a neural level. For the near future, this approach will be useful and will provide interesting insights, and the cognitive model-imaging combination can, when successful, generate a set of target functions at which neurally based large-scale modeling can aim. It is to these bottom-up, biologically based models that we now turn.

3 Bottom-up Methods

The "bottom-up" type of neural modeling has variously been referred to as synthetic functional brain imaging (Arbib et al. 1995), or large-scale neural modeling (Horwitz & Tagamets 1999; Husain et al. 2004; Tagamets & Horwitz 1998), or forward or generative modeling (David et al. 2005). An important goal is to relate neural electrical activity to functional neuroimaging data. This simulation approach is in many ways more ambitious, and less (PET/fMRI) data driven, than is the data-fitting use of neural modeling, but is crucial for furthering our understanding of the neural basis of behavior. Importantly, it entails determining both the neural basis for local brain activations and the neurobiological correlates for the PET/fMRI-determined functional connections.

We will illustrate this type of modeling by examining a number of studies that addressed different questions associated with the relationship between neural activity on one hand and functional neuroimaging on the other. The first set of studies focuses on the relation between neural activity and the corresponding hemodynamic response that is measured by PET and fMRI. The second set illustrates how this approach enables one to integrate neural information across different spatiotemporal scales. The third subsection discusses the use of large-scale neural modeling to help understand the neural bases of functional and effective connectivity.

3.1 Excitatory and Inhibitory Neural Activity

The first paper that discussed ways to relate neural modeling to functional brain imaging data was by Horwitz and Sporns (1994). The first actual study that compared simulated data generated by a biologically-based neural model to hemodynamic-based functional neuroimaging data was by Arbib et al. (1995), who used a large-scale model of saccade generation (Dominey & Arbib 1992) and adapted it to generate simulated PET data. Their model included a number of brain structures, such as posterior parietal and visual cortex, superior colliculus, the frontal eye field, the mediodorsal and lateral geniculate nuclei of the thalamus, and the caudate nucleus and substantia nigra of the basal ganglia. Because some of the pathways involving the basal ganglia are inhibitory, this model was a good testing ground for examining the effects

of inhibitory synaptic activity on simulated blood flow. The main hypothesis tested in the PET simulation was that regional cerebral blood flow, as measured by PET, correlates with local synaptic activity. Although at the time there was some, but not very much, experimental support for this hypothesis (Jueptner & Weiller 1995), since the publication of the Arbib et al. study, the evidence has grown much stronger that the hemodynamic methods are indicative of synaptic and postsynaptic activity (e.g., Lauritzen 2001; Logothetis et al. 2001). One consequence of this notion is that increases in excitatory and inhibitory synaptic activity can lead to increased blood flow and metabolic activity (Logothetis 2003). In the Arbib et al. (1995) study, PET activation in the model was computed by summing the absolute values of both excitatory and inhibitory synaptic weights times firing rates of presynaptic neurons and integrating these values over a time period that corresponded to the time scale of PET while the model performed a specific task.

Computed PET activity was calculated by Arbib et al. (1995) during two different tasks (generating simple saccades and memory-driven saccades) and the differences between the two conditions were evaluated in all regions of the model. Memory driven saccades are generated in the model by activation of a memory loop between the frontal eye field and mediodorsal nucleus (MD) of the thalamus, which is disinhibited by the substantia nigra, generating a saccade (via the superior colliculus) to a remembered target when there is no stimulus present. When compared to the simple saccade task, in which disinhibition of the superior colliculus allows a saccade to be generated to a target present in the field of view, spiking activity in the modeled MD region increased as a result of the disinhibition, while simulated PET activation decreased in MD. This result showed how modeling can illuminate a counterintuitive effect: during the simple saccade, synaptic activity from the tonic inhibition of the MD contributes more synaptic activity to the PET measure than the increase in excitation that results from disinhibition.

As shown by the above study, the interpretive difficulty associated with synaptic inhibition provided a good example of how large-scale modeling can help interpret the activations observed in functional neuroimaging studies, even when they lead to results that appear counterintuitive. Because different parts of the brain have different neural architectures, and because the composition of the excitatory and inhibitory elements will be different in these various architectures, a number of separate modeling efforts will be needed to understand fully the role of inhibition in PET/fMRI activation patterns. For neocortex the inhibition theme was explored by Tagamets and Horwitz (2001) using a large-scale model of the ventral cortical visual processing stream (the details of this model will be discussed below). They used the same assumption as did Arbib et al. (1995) – that PET regional cerebral blood flow is indexed by the absolute value of the total synaptic activity in a brain region. They identified three factors that may play a role in how neural inhibition affects imaging results: (1) local connectivity; (2) context (e.g., the type of task being performed by the network); and (3) type of inhibitory connection.

Simulation results showed how the interaction among these three factors can explain seemingly contradictory experimental results. Specifically, the modeling indicated that neuronal inhibition can raise brain imaging measures if there is either low local excitatory recurrence or if the region is not otherwise being driven by excitation. On the other hand, with high recurrence or actively driven excitation, inhibition can lower observed neuroimaging values.

To summarize this section, we have shown that an important use for large-scale modeling is to help interpret how task-specific mixtures of excitatory and inhibitory neural activities result in the complex patterns of activations often seen in functional neuroimaging studies.

3.2 Integrating Neuroscientific Data Across Spatiotemporal Scales

We have already emphasized that the major goal of large-scale neural modeling is to enable one to propose a set of neural-based mechanisms that can explain particular human cognitive functions. If these mechanisms are reasonable representations of what actually occur in the brain, then simulated data generated by the model, *at multiple spatiotemporal scales*, should closely approximate corresponding experimental data. We will illustrate this use of modeling by discussing a model of object processing that was developed in our laboratory.

There are two versions of this model, one for processing visual objects (Tagamets & Horwitz 1998) and one for auditory objects (Husain et al. 2004). Although the notion of visual object (e.g., chair, table, person) seems straightforward, that of auditory object is more elusive, but can be thought of as an auditory perceptual entity subject to figure-ground separation (for a detailed discussion see Griffiths & Warren 2004; Kubovy & Van Valkenburg 2001). Examples of auditory objects would include words, melodic fragments, and short environmental sounds. There is much experimental evidence implicating the ventral visual processing pathway, which runs from primary visual cortex in the occipital lobe into the inferior temporal lobe and thence to inferior frontal cortex, as being involved with visual object processing in human and nonhuman primates (Desimone & Ungerleider 1989; Haxby et al. 1994; Ungerleider & Mishkin 1982). Although the supporting evidence is less extensive, an analogous processing stream along the superior temporal gyrus (STG) for auditory object processing has been hypothesized by Kaas, Rauschecker and others (Kaas et al. 1999; Rauschecker & Tian 2000). Our models build on these notions; we have proposed (and instantiated in our models) that visual and auditory (and possibly tactile) object processing uses a set of similar cortical computational mechanisms along each of their respective pathways, although the features on which these mechanisms act depend on the sensory modality (Husain et al. 2004). However, it is important to emphasize that we are not implying that all sensory features have analogues in the three systems, only that some do.

Visual model

The visual model (Tagamets & Horwitz 1998) performs a set of tasks, specifically a delayed match-to-sample (DMS) task for shape, and a "passive" viewing task. We chose the DMS task because there exist much functional neuroimaging, neuroanatomical, electrophysiological, and cognitive data in human and nonhuman primates relevant for this type of task. The DMS task involves the presentation of a shape, a delay, and the presentation of a second shape; the model has to decide if the second stimulus is the same as the first. Multiple trials (e.g., 10) are used to simulate an fMRI or PET study.

We incorporated four major brain regions of the ventral occipitotemporal pathway into the visual model: ((1) primary visual cortex (V1/V2); (2) occipitotemporal cortex (V4); (3) inferior temporal cortex (IT); and (4) prefrontal cortex (PFC) (see Fig. 1). Each region contains populations of neuronal

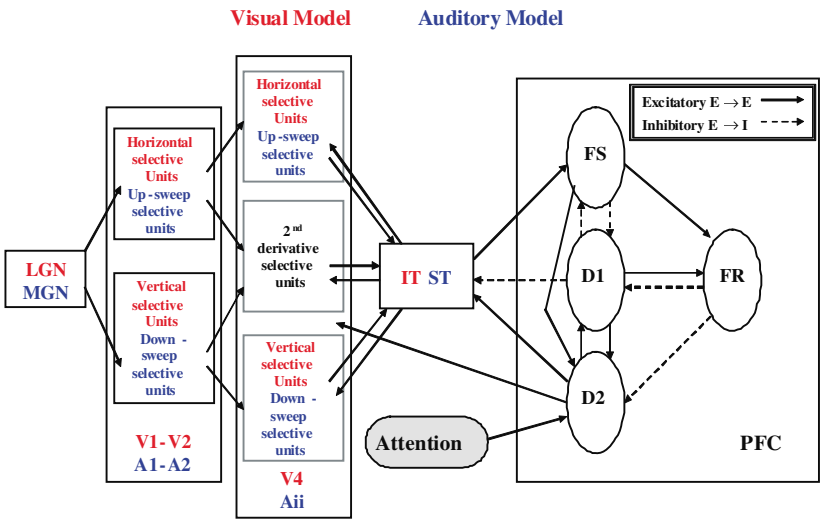


Fig. 1. Network diagram of the object processing model (Tagamets & Horwitz 1998; Husain et al. 2004). The regions of each model form a complex network of feedforward and feedback connections; these interregional connections can be either excitatory (excitatory-to-excitatory elements, shown as solid lines) or inhibitory (excitatory-to-inhibitory elements, shown as dashed lines). Regions specific to the visual model (LGN, V1–V2, V4, IT) are shown in red; regions specific to the auditory model (MGN, A1–A2, Aii, ST) are shown in blue; the prefrontal module is structured in the same way in both models. The sub-modules for each model are also indicated. In the PFC region, there are four submodules: FS contains stimulus-sensitive units, D1 and D2 contain units active during the delay part of a delayed match-to-sample task, and FR contains units whose activity increases if there is a match between the first and second stimuli of a trial. Abbreviations: LGN – lateral geniculate nucleus; MGN – medial geniculate nucleus; Aii - secondary sensory auditory cortex; PFC – prefrontal cortex

assemblies of basic units, each of which is an interacting excitatory-inhibitory leaky integrator neuronal pair that represents a simplified cortical column (Wilson & Cowan 1972). Although there are many attributes that can be used to characterize an object (e.g., shape, color, texture), we chose to focus on shape and assumed that the basic feature, based on the work of Hubel and Wiesel (Hubel & Wiesel 1977), is line orientation. So, the excitatory neurons in the V1/V2 module were constructed to respond in a retinotopically configured manner to lines oriented in particular directions (for computational simplicity, we use only horizontal and vertical lines). The V4 module is similar to the V1/V2 one in that it is retinotopically organized, and contains neurons with horizontal and vertical line orientation selectivity. It also contains neurons that respond to a combination of features (i.e., neurons that respond best to a change in line orientation, what can be called second-derivative neurons). Importantly, the receptive field of these neurons, like those of actual V4 neurons, are larger than those for V1/V2 neurons (by about a factor of 3). This was achieved by increasing both the divergence and sparseness of connections in the feedforward direction, with neighboring V1/V2 units sending connections to neighboring V4 units. Likewise, V4 units project to IT with the same 3-fold divergence, with the result that IT units no longer show retinotopy and each IT neuron's receptive field is essentially the entire visual field, which is the situation with real IT neurons (Desimone et al. 1984). The upshot of this is that a visual object is represented in IT in a highly distributive manner. In the PFC module, we have four submodules whose neuronal units have the response properties based on the findings of Funahashi et al. (1990): units that respond when a visual stimulus is present, two kinds of units that show activity during the delay interval, and units whose activities increase when a match between the second and first stimuli occurs. A particular arrangement of connectivity between these four submodules, and between these submodules and the other regions, enable a short-term memory trace to be maintained during the delay interval between the presentation of the two stimuli in the DMS task.

Feedforward and feedback connections between regions were based, where available, on primate neuroanatomical data. Parameters were chosen so that the excitatory neuronal elements of each module had simulated neuronal activities resembling those found in electrophysiological recordings from monkeys performing similar tasks (e.g., Funahashi et al. 1990).

The method by which the "task instructions" are implemented, so that the model knows whether to perform the DMS task or a control task, is by means of a continuous "attention" or "biasing" variable that modulates a subset of prefrontal units by means of diffuse synaptic inputs. These prefrontal units, through feedback connections, modulate activity in posterior areas. The strength of the biasing variable controls whether the stimuli are maintained in short-term memory or not. As a consequence, activity in each brain area is a mixture of feedforward activity ("bottom-up") determined in part by the presence of an input stimulus, feedback activity ("top-down") determined in

part by the strength of the biasing attention signal, and local activity within each region.

Functional neuroimaging studies are simulated by presenting stimuli to an area of the model representing the lateral geniculate nucleus (LGN) for the visual model. Approximately 10 such trials per condition would be used for a PET or fMRI study. To simulate the data acquired in a PET scan, we integrated the absolute value of the summed synaptic activities in each brain region over all the trials, separately, for the DMS condition and for a control condition ('passive' viewing of degraded shapes) (Tagamets & Horwitz 1998). In the case of fMRI, regional activity is simulated by first temporally and spatially integrating the absolute value of the synaptic activity in each region over a time period representing 50–100 m sec, which corresponds to the time needed to collect a single slice of fMRI data. The time integrated synaptic activities are then convolved with a function representing hemodynamic delay, and subsequently sampled at a time corresponding to the repetition time (TR) of an fMRI scan (Horwitz & Tagamets 1999).

The simulated PET values for the visual model of the DMS condition, when compared to the control condition, were similar (Tagamets & Horwitz 1998) to those found in experimental PET studies of face working memory (Haxby et al. 1995), as shown in the bar graph on the left of Fig. 2. In

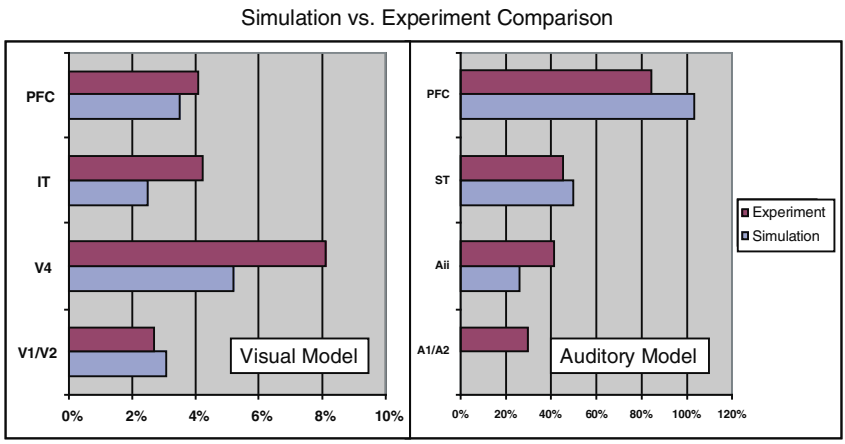


Fig. 2. Comparison between model simulation and experimental results. (Left) Bar graphs of the percent signal change in simulations using the visual model (blue bars) in each brain region and the corresponding results (red bars) from the comparable PET study of (Haxby et al. 1995). The percent signal changes were evaluated as the difference between the DMS task of object shape vs. a passive viewing task of degraded shapes (see Tagamets & Horwitz (1998) for details). (Right) A similar comparison for the auditory model and the corresponding fMRI study. The percent signal changes were computed as the normalized difference of the percent signal changes of both the DMS and control tasks relative to baseline (see Husain et al. (2004) for details)

summary, the visual model was able to generate simulated electrophysiological data and simulated PET data that generally were in close agreement with experimental data. Moreover, the model could perform the DMS task.

We have dwelt at length on the visual model because many of the assumptions used in its construction have significant experimental support, and consequently, require little justification on our part. The really new thing is the way in which these multiple mechanisms are put together. What the modeling shows is (1) these quantitative agreements between simulation and experiment strongly argue for the specific hypotheses that were made concerning the neural mechanisms by which multiple interacting brain regions implement this visual DMS task; (2) the assumptions used to related the neural activity to the functional neuroimaging data are sound; and (3) it is possible to account for neuroscientific data at a macroscopic (brain area) level in terms of activity at a mesoscopic (cellular/columnar) level.

Auditory Model

The large-scale auditory model that was developed by Husain et al. (2004) implements a DMS task for auditory objects. The key features that we chose to model were frequency sweeps, and the auditory objects, which we call tonal contours (TC), consist of combinations of sweeps and pure tones, each TC lasting about 350 m sec (Fig. 3). Other auditory features, such as timbre, are not included in the model. Like the visual model, the auditory model contains four major modules (see Fig. 1): (1) primary sensory cortex (A1/A2), where neurons respond in a tonotopically-organized fashion to up and down frequency sweeps ; (2) secondary sensory cortex (e.g., lateral and parabelt auditory cortex, labeled Aii), where neurons respond to longer up and down sweeps, as well as to changes in the direction of frequency sweeps; (3) an area in the superior temporal gyrus and/or sulcus (ST) similar to IT where a complex sound is represented in a distributive manner in the neuronal population; and (4) a prefrontal module analogous to that in the visual model.

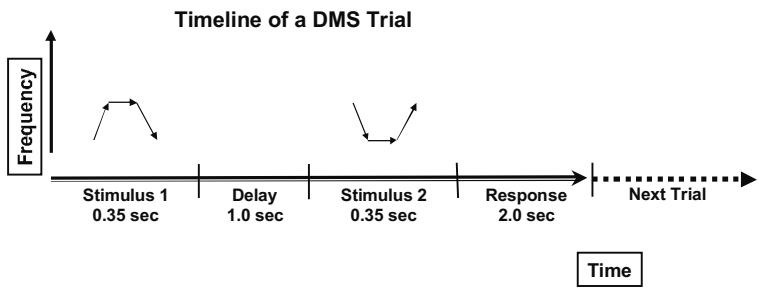


Fig. 3. Timeline of a single delayed match-to-sample (DMS) trial for the auditory model (and the corresponding auditory fMRI experiment). Shown are frequency vs. time representations of the tonal contours (TCs)

The experimental evidence for the presence of neurons with these properties was reviewed in Husain et al. (2004). In brief, there were reports showing that there are neurons in auditory cortex that respond to the direction of frequency sweeps (e.g., Mendelson & Cynader 1985), and that there are neurons in prefrontal cortex that are active during the delay interval in a delayed response task for auditory stimuli (Kikuchi-Yorioka & Sawaguchi 2000). However, unlike the situation in visual cortex, there have been relatively few studies in awake monkeys and other mammalian preparations in which the response properties of neurons in various parts of the STG were evaluated. So, for example, there is only a small amount of evidence showing that neurons in the anterior part of the STG respond to complex auditory patterns (e.g., Kikuchi et al. 2004). Similarly, a number of other crucial assumptions that Husain et al. (2004) made in constructing the auditory model also rested on either weak experimental data (i.e., just a few studies), or else were made in analogy to the visual case. To illustrate, a key assumption we used was that in going from primary to secondary auditory cortex and thence to ST, the spectrotemporal window of integration increased (analogous to the increase in the size of the spatial receptive field in the visual model), so that, for example, neurons in the secondary auditory area respond best to longer frequency sweeps that do the neurons in A1/A2. Experimental support for this assumption could only be found in Harrison et al. (2000). Another key assumption was that in the secondary auditory area there was a neuronal population that responded best to a change in the direction of frequency sweeps. There was no published evidence for such 'second-derivative' neurons in auditory cortex.

In essence, each of these assumptions can now be considered predictions of the model. To focus on a case that we shall come back to, the assumption concerning the increase in the spectrotemporal window of integration is instantiated in the model by increasing both the divergence and sparseness of connections in the feedforward direction, with neighboring A1/A2 units sending connections to neighboring units in secondary auditory cortex. Likewise, these latter units project in a similar divergent fashion to ST. Consequently, the assumption of increasing spectrotemporal receptive field size ultimately rests on assumptions concerning the pattern of interregional neuroanatomical connectivity in STG.

Identical stimuli were used for both modeling and the corresponding fMRI experiment (Husain et al. 2004). The bar graph on the right of Fig. 2 shows the percent signal changes (comparing the fMRI activity for tonal contours to that for pure tones) in each brain region for both the simulated and the experimental data. Two important points arise from this part of Fig. 2. First, our simulated results in primary auditory cortex (A1/A2) did not match the experimental value (in the simulation, the percent change between tonal contours and tones was near zero). A likely reason for this is that we included in our model only one type of neuron found in primary auditory cortex (selective for frequency sweeps), but there are many neuronal types in the brain selective for other features in the auditory input (e.g., loudness, on and off properties

of the input) that we were not modeling. Moreover, there was a large amount of scanner noise during the experiment that could have had some effect on the experimental data that was not taken into account in the simulation. The second important point is that we were able to get close quantitative agreement between simulated and experimental data in all the right hemisphere regions that corresponded to those in the model (except, as noted above, A1/A2). As far as we know, this was the first study in which a biologically realistic neural model generated stimulated fMRI data that generally agreed *quantitatively* with experimental fMRI values in which task design and stimuli were identical to those used in the modeling.

To test the robustness of the auditory model, we (Husain et al. 2005) used it to investigate the auditory continuity illusion, which is an example of one type of auditory perceptual grouping phenomenon. Perceptual grouping permits the auditory system to integrate brief, disparate sounds into cohesive perceptual units, which is important for perception because it enables, for example, one to separate attended sounds from environmental noise. The auditory continuity illusion emerges when a sound object (e.g., pure tone, frequency sweep, word) is perceived to continue through occluding noise even though no such signal need be physically present in the noise. Although it serves the important purpose of making communication sounds intelligible in a noisy environment and although it been extensively studied by means of psychophysical experiments, little is known concerning neural basis of this illusion.

In our simulations, intact stimuli (tonal contours) were matched with fragmented versions (i.e. with inserted silent gaps) of the stimuli (Fig. 4). The ability of the model to match fragmented stimuli declined as the duration of the gaps increased (Fig. 5, top). However, when simulated broadband noise was inserted into these gaps, the matching response was restored indicating that a continuous stimulus was perceived (Fig. 5, bottom). The electrical activities of the neuronal units of the model agreed with electrophysiological data obtained by Sugita (1997), and the behavioral activity of the model matched human behavioral data (Ciocca & Bregman 1987; Dannenbring 1976). The most important aspect of this study relevant to this chapter concerns how our model implements the illusion. The predominant mechanism is the divergence of the feedforward anatomical connections along the auditory processing pathway in the temporal cortex. Not only do our results attest to the robustness of the model, but further, they predict the primary role of the anatomical connectivity of the auditory processing areas in mediating the continuity illusion. Note that these results were obtained without changing any of the parameters of the auditory model.

In summary, our simulation results for the auditory model demonstrate that our assumptions concerning the neural mechanisms by which auditory objects are processed in the cerebral cortex are such that they enabled us (1) to simulate neuronal data from multiple brain regions that agreed with available experimental data, (2) to simulate fMRI data that quantitatively

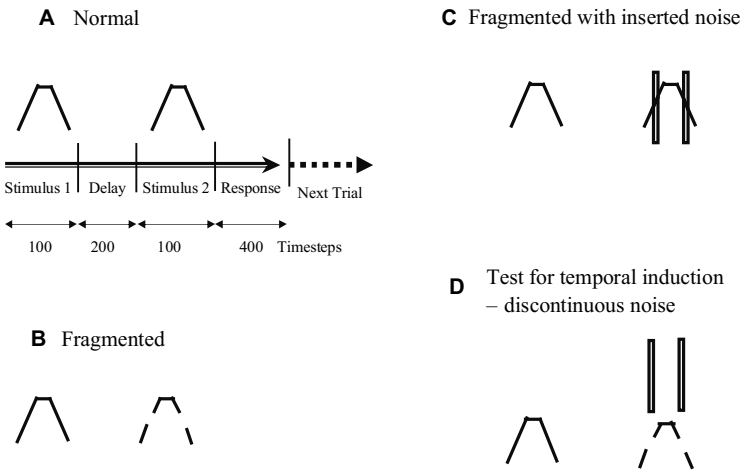


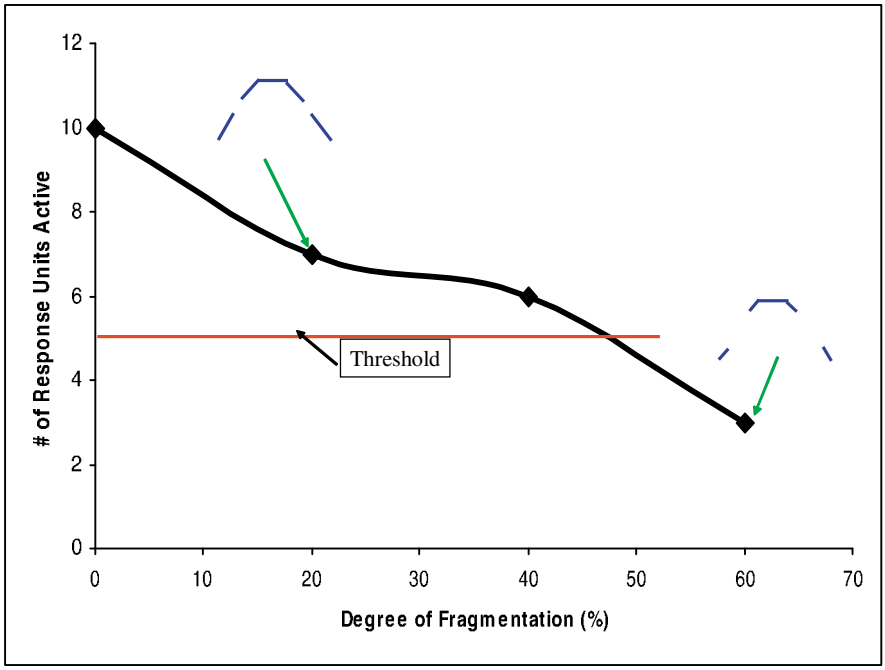
Fig. 4. Illustration of the auditory continuity illusion. (**A**) Shown is the DMS task for a tonal contour (indicated as a frequency vs time pattern; each tonal contour is about 350 m sec in duration). (**B** & **C**) Fragmented tonal contours, with and without inserted noise. If the noise spans the frequency range of the contour, and is loud enough, the fragmented tonal contour is perceived as intact. (**D**) If the noise is outside the range of the contour, the fragmented tonal contour is not perceived as intact. Modified from Husain et al. (2005)

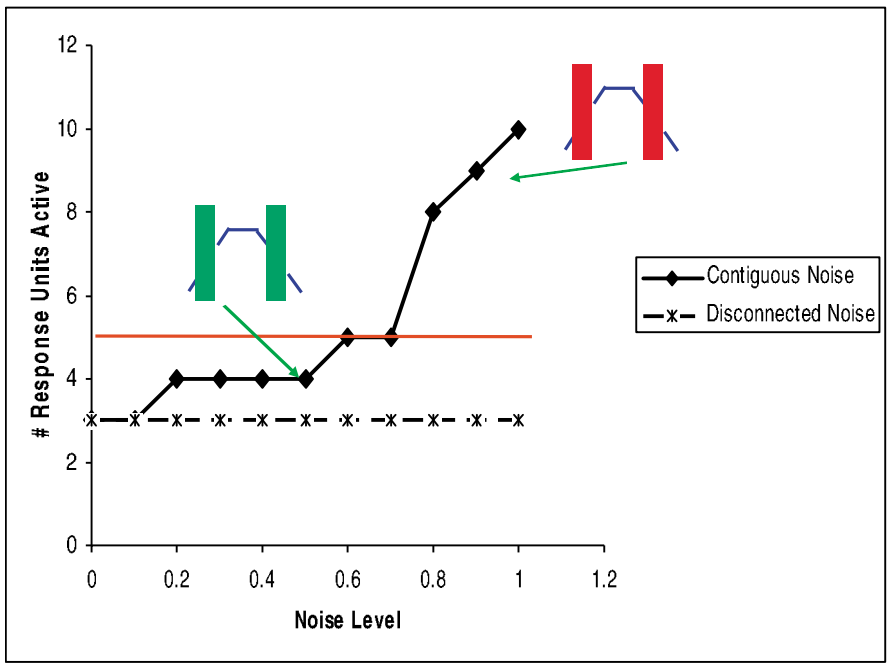
matched experimental data, again in multiple brain regions, and (3) to generate simulated behavioral data for the auditory continuity illusion that matched experimental data obtained in humans.

Other Large-Scale Modeling Work

There have been several other recent studies in which large-scale neural modeling has been used to relate neural activity to functional neuroimaging data. We will briefly mention a few of them.

Deco et al. (2004) modeled the mechanisms that underlie working memory-related activity during the execution of delay tasks that have a "what"-then-"where" design (with both object and spatial delayed responses within the same trial). They were interested in examining two notions related to the topographical and functional organization of the PFC: (1) organization-by-stimulus-domain, which proposes that dorsolateral PFC is involved with spatial processing and ventrolateral PFC is specialized for object processing (e.g., Wilson et al. 1993); (2) organization-by-process, which puts forward a hierarchical organization of the PFC, with non-memory related higher order functions (e.g., manipulation of items in memory) associated with dorsolateral PFC regions, and short-term memory maintenance functions ascribed to ventral PFC (e.g., Petrides 1994). Deco and colleagues utilized a network composed of integrate-and-fire neurons to model both single-neuron and fMRI





data on short-term memory in topologically different parts of the PFC during delay tasks having the what-then-where design. The model contained different populations of neurons (as found experimentally) in attractor networks that responded in the delay period to the stimulus object, the stimulus position, and to combinations of both object and position information. These neuronal populations were arranged hierarchically and global inhibition mediated through inhibitory interneurons was employed to implement competition. The relative activity of the different attractor populations required to perform what-then-where and where-then-what short-term memory tasks was provided by an external attentional signal that biases the different neuron populations within the framework of the biased competition model of attention (Desimone & Duncan 1995; Rolls & Deco 2002).

It was shown that their model could account for the neurophysiological activity seen in both the ventrolateral and dorsolateral PFC during the delay periods of working memory tasks utilizing the what-then-where design, obtained by Rao et al. (1997). Furthermore, the Deco et al. model generated simulated fMRI patterns that matched experimental findings during a what-then-where short-term memory task for both PFC sectors as shown by the fMRI findings of Postle & D'Esposito (1999). However, this could not be done if it was assumed that the difference between ventrolateral and dorsolateral PFC followed the organization-by-stimulus-domain hypothesis, with the dorsal PFC being specialized for spatial processing and ventral PFC being specialized for object processing. Rather, Deco et al. (2004) had to assume that the differences between these two prefrontal regions arose from having a larger amount of inhibition in the dorsolateral portion of the PFC than in the ventrolateral part.

The Deco et al (2004) study is important for several reasons. First, it is the first large-scale neural model applied to fMRI data that employed spiking neurons. Second, it demonstrates how this type of modeling can shift the terms of a scientific debate. There is great deal of controversy concerning the different functional roles of dorsal and ventral PFC. Heretofore, the controversy was cast in cognitive terms (i.e., spatial vs. object processing on the one hand,

Fig. 5. Performance of the auditory model of Husain et al. (2004) for the auditory continuity illusion. Above-threshold activity of 5 or more neurons in the response module (FR) indicates a match between the two stimuli of a DMS trial. The top graph shows that with a short duration gap in the second stimulus, the model (like actual subjects) indicates a match, thus grouping the parts of the tonal contour into a perceptual whole. As the gap widens, a non-match results. If noise is inserted in the gap, and is of weak intensity (green), the tonal contour is not considered as continuing through the gap (bottom); if the noise is more intense (red), then perceptual grouping occurs, but only if the band of noise is in the part of frequency space occupied by the tonal contour. See Husain et al. (2005) for details. Modified from Horwitz & Glabus (2005)

manipulation vs. maintenance on the other). The results of the Deco et al. study demonstrate that one can rephrase the debate in neural terms (i.e., different neuronal populations compared to different levels of inhibition), and in these terms, a variety of neuroscientific results can be brought forward to resolve the issue.

The next study we present comes from Chadderon and Sporns (in press) and it also focuses on prefrontal cortex. A large-scale computational model of prefrontal cortex and associated brain regions was constructed. It was designed to investigate the mechanisms by which working memory and task state interact to select adaptive behaviors from a behavioral repertoire. The model consisted of multiple brain regions containing neuronal populations with realistic physiological and anatomical properties: extrastriate visual cortical regions, inferotemporal cortex, prefrontal cortex, striatum, and midbrain dopamine neurons. Like the visual model of Tagamets and Horwitz (1998) discussed earlier, the Chadderon-Sporns model used Wilson-Cowan leaky integrator neurons (Wilson & Cowan 1972).

In the Chadderon-Sporns model the onset of a delayed match-to-sample or delayed non-match-to-sample task triggers tonic dopamine release in prefrontal cortex, which causes a switch into a persistent, stimulus-insensitive dynamic state that promotes the maintenance of stimulus representations within prefrontal networks. Other modeled prefrontal and striatal units select cognitive acceptance or rejection behaviors according to which task is active and whether prefrontal working memory representations match the current stimulus. Working memory task performance and memory fields of prefrontal delay units were degraded by extreme elevation or depletion of tonic dopamine levels. Analyses of cellular and synaptic simulated activity indicated that hyponormal dopamine levels resulted in increased prefrontal activation, whereas hyper-normal dopamine levels led to decreased activation.

Chadderon and Sporns also used their simulated results to derive synthetic fMRI signals, in a similar manner to that discussed earlier (Horwitz & Tagamets 1999). They found that under normal dopamine conditions, there was a significant increase in PFC fMRI activity in a DMS working memory task, as compared to an "idle" control condition. If a relatively fast hemodynamic delay function was used to derive fMRI signals, the increase was confined to the delay periods of the task, and absent during the cue/distractor/target periods. Decreasing tonic dopamine levels led to higher baseline activation of PFC, but the activity differences between idle and working memory conditions were not significant. Conversely, if tonic dopamine levels were elevated, baseline activation of PFC was reduced, and activity during a working memory task was decreased with respect to corresponding control trials, thus showing a reversal of the increase found under normal dopamine levels.

The Chadderon and Sporns model (Chadderdon & Sporns in press) represents an important step forward over the previous models we discussed in that it explicitly incorporates a modulatory neurotransmitter (dopamine) so that more complex behavior can be addressed. In particular, they implemented

two separate domains of prefrontal working memory. Task identity (i.e., delayed match-to-sample, delayed nonmatch-to-sample and their corresponding control tasks) was maintained by a segregated set of recurrently excitatory and mutually inhibitory cell populations. Stimulus feature memory was maintained by tonic dopamine level. It seems likely that in the near future, the level of model complexity will increase even further.

The final example we present in this subsection involves a model that simulates EEG/MEG dynamics. As mentioned in the introductory section, EEG/MEG activity has engendered a number of computational neural modeling efforts (e.g., David & Friston 2003; Jirsa & Haken 1997; May et al. 1999; Nunez 1981; Robinson et al. 2005). Here, we shall discuss a recent study by David et al. (2005) that describes a neurally plausible forward model designed to reproduce responses seen empirically, thus allowing mechanistic enquiries into the generation of evoked and induced EEG/MEG responses to be made. As these authors point out, most neural models of EEG/MEG were constructed to generate alpha rhythms (e.g., Jansen & Rit 1995; Stam et al. 1999), but recent work has shown that models that produce the entire spectrum of EEG/MEG oscillations can be created (e.g., David & Friston 2003; Robinson et al. 2001). The model developed by David and colleagues (David et al. 2005) focuses on simulating event-related activity.

One important aspect of the David et al. model is that the basic functional unit, based on the work of Jansen and Rit (1995), uses three neuronal subpopulations to represent a cortical area; one subpopulation corresponds to pyramidal neurons, a second represents excitatory interneurons and the third corresponds to inhibitory interneurons. A second important feature is that their neural mass model consists of hierarchically arranged areas using three kinds of inter-area connections (forward, backward and lateral). The excitatory interneurons can be considered to be spiny stellate cells found primarily in layer 4 of the cortex; they receive forward connections. The excitatory pyramidal neurons are the output cells of a cortical column, and are found in the agranular layers, as are the inhibitory interneurons. The MEG/EEG signal is taken to be a linear mixture of the averaged depolarization of the pyramidal neurons.

Using this model, David et al. (2005) investigated how responses, at each level of a cortical hierarchy, depended on the strength of connections. They did this in the context of deterministic responses and then with stochastic spontaneous activity. One important result of their simulations was that with the presence of spontaneous activity, evoked responses could arise from two distinct mechanisms: (1) for low levels of (stimulus related and ongoing) activity, the systems response conforms to a quasi-linear superposition of separable responses to the fixed and stochastic inputs, which is consistent with the traditional assumptions that motivate trial averaging to suppress spontaneous activity and reveal the event-related response; (2) when activity is sufficiently high, there are nonlinear interactions between the fixed and stochastic inputs, which results in a phase resetting, which in turn leads to a different explanation for the appearance of an evoked response.

3.3 Investigating the Neural Substrates of Functional and Effective Connectivity

In the last section, we presented a number of examples in which large-scale neural models were used to simulate PET and fMRI data. In both the simulated and experimental cases, the data of interest usually are the "activations", which are the differences between regional activities in the condition of interest compared to activities in a control condition. However, a second, and increasingly important, way to analyze functional neuroimaging data is to use the covariance paradigm (Horwitz et al. 1992), which asserts that a task is mediated by a network of interacting brain regions, and that different tasks utilize different functional networks (see also McIntosh 2000). The fundamental concept employed by the covariance paradigm is *functional connectivity*, which is used in the context of neuronal processes to allude to the functional interactions between different areas of the brain. The formal mathematical quantification of these functional relationships depends on the use of interregional covariance or correlation coefficients (e.g., Friston 1994). However, the actual definition used by different researchers varies widely, as do the computational algorithms employed to evaluate interregional functional connectivity (Horwitz, 2003).

Functional connectivity doesn't distinguish explicitly between the case where regions are directly influencing one another (e.g., along specific anatomical pathways) compared to simply indicating some kind of indirect interaction. To do the former, some type of computational top-down modeling needs to be employed (Friston 1994; Horwitz 1994; McIntosh & Gonzalez-Lima 1994) to calculate the interregional *effective connectivity*. This term has come to mean, at the systems level, the direct influence of one neural population on another (Friston 1994). We think of it as meaning the functional strength of a particular anatomical connection; for instance, during one task a specific anatomical link may have a stronger effect than during a second condition; likewise, such a link may be stronger in normal subjects than in a patient group. The evaluation of effective connectivity requires modeling because one needs to select a small group of brain regions to include in the network, and one needs to combine functional neuroimaging data with information about the anatomical linkages between these regions. That is, the validity of effective connectivity modeling builds on a combination of implied anatomical and functional connections between brain regions. A number of chapters in this volume address functional and effective connectivity; see, for example, the contributions by Stephan and Friston, by Bullmore, and by Strother.

As we have emphasized in several previous papers (Horwitz 2003; Horwitz et al. 2005), there are multiple measures of functional and effective connectivity in use, and there is no guarantee that the conclusions one draws using one measure will be the same as using another. Furthermore, the terms functional and effective connectivity are applied to quantities computed on types of functional imaging data (e.g., PET, fMRI, EEG) that vary in spatial, temporal,

and other dimensions, using different definitions (even for data of the same modality) and employing different computational algorithms. Until it is understood what each definition means in terms of an underlying neural substrate, comparisons of functional and/or effective connectivity across studies may lead to inconsistent or misleading conclusions. Perhaps more important is the fact that since the neural substrates of each measure are unknown, it is unclear how well a particular way in which the functional or effective connectivity is computed accurately represents the underlying relationships between different brain regions.

To address these issues, we have started using our large-scale neural models to help determine the neurobiological substrates for the most widely used definitions and algorithms for evaluating interregional functional and effective connectivity. That is, we simulate fMRI time series with our models and compute a particular version of the functional or effective connectivity. Because in the model we know what each neuron and synapse are doing at every moment in time (unlike the situation for real brain data), we can determine how well the computed functional/effective connectivity reflects the actual underlying neural relationships in the models. Because our models are neurobiologically plausible, complex, contain both excitatory and inhibitory neurons, have feedforward and feedback connections and include a diversity of regions containing neurons that possess different response properties, they provide useful testing grounds for investigating various kinds of data analysis methods (Horwitz 2004).

One example of this approach can be found in Horwitz et al. (2005), where one of the simplest definitions of fMRI functional connectivity – the within-condition correlation between fMRI time series - was examined. The crucial aspect of simulating functional and effective connectivity is to be able to simulate in biologically plausible ways variability in neural activity, because the key to evaluating functional or effective connectivity is to assess interregional co-variability. There are multiple sources of the variability found in functional neuroimaging data. Although some of these originate from the scanning technique and some are non-neural in origin (e.g., changes in the vasculature may lead to changes in the fMRI hemodynamic response function), some of the variability observed in the functional neuroimaging signal can be utilized to provide the covariance needed to evaluate functional or effective connectivity. The main idea is that variability in the activity in one region of the neural network mediating the task under study is propagated to other regions, resulting in a larger covariance between the regions than would be the case if they were not interacting with one another. The various methods of evaluating functional connectivity try to tap one or more of these neurally-based sources of covariation. For fMRI, the three main sources of variation that can be employed to assess within-condition functional connectivity are subject-to-subject variability, block-to-block variability and item-to-item (or MR volume-to-volume) variability (see Horwitz et al. (2005) for a more extensive discussion of this topic).

In our modeling approach, these types of variability were incorporated as subject-to-subject differences in the strengths of anatomical connections, scan-to-scan changes in the level of attention, and trial-to-trial interactions between the neurons mediating the task and nonspecific neurons processing noise stimuli (Horwitz et al. 2005). Recall that in our modeling framework, simulated fMRI is computed by integrating the absolute value of the synaptic activities in each module every 50msec (we call this the integrated synaptic activity or ISA), convolving this with a function representing the hemodynamic response and then sampling the resulting time series every TR seconds. Because the hemodynamic convolution and sampling lead to a loss of temporal resolution, one can think of the functional connectivities calculated from the ISA as a kind of 'gold standard' in that they represent the most ideal evaluation of the neural interrelationships that one could get at the systems level. Indeed, it has been argued by Gitelman and colleagues (Gitelman et al. 2003) that deconvolving an experimental fMRI time series improves the evaluation of functional and effective connectivity.

Using the visual model of Tagamets and Horwitz (1998), we (Horwitz et al. 2005) explored the functional connectivity, evaluated as the within-condition correlation between fMRI time series. Focusing on the link between IT and PFC (see Fig. 1 above), we found that time series correlations between ISAs between these two modules were larger during the DMS task than during a control task. These results were less clear when the integrated synaptic activities were hemodynamically convolved to generate simulated fMRI activities. In a second simulation using the auditory model of Husain et al. (2004), we found that as the strength of the model anatomical connectivity between temporal and frontal cortex was weakened, so too was the strength of the corresponding functional connectivity, although the relation was nonlinear. This latter result is important, since it demonstrates that the fMRI functional connectivity can appropriately reflect the strength of an anatomical link.

A final illustration of using large-scale neural modeling to investigate the neural substrates of functional/effective connectivity is a recent study by Lee et al. (in press). Friston and colleagues (Friston et al. 2003) developed a method called Dynamic Causal Modeling (DCM) for estimating and making inferences about the changes in the effective connectivities among small numbers of brain areas, and the influence of experimental manipulations on these couplings (see the article by Stephan and Friston in this volume for a thorough discussion of this method). Lee et al. used DCM to evaluate the change in effective connectivity using simulated data generated by our large-scale visual model implementing the DMS task. System-level models with hierarchical connectivity and reciprocal connections were examined using DCM and Bayesian model comparison (Penny et al. 2004), and revealed strong evidence for those models with correctly specified anatomical connectivity. An example of this is illustrated in Fig. 6. A simple model incorporating the first three regions of the large-scale visual model (V1/V2, V4 and IT) was analyzed. Three arrangements of the anatomical linkages between regions were compared: (1) the

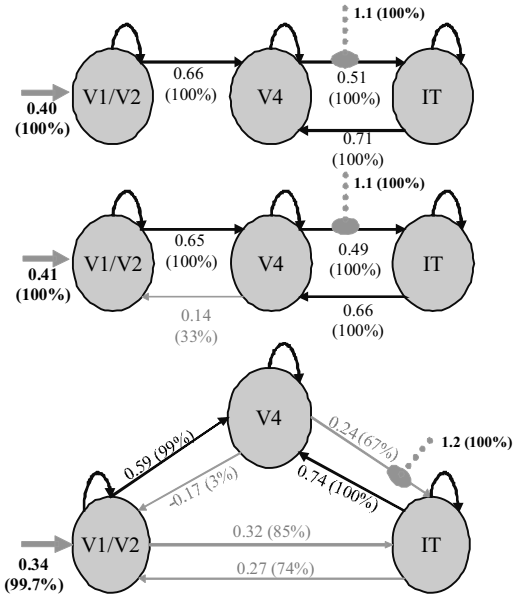


Fig. 6. DCM analysis of visual model. (Top) Coupling parameters for the simple model, with interarea connectivity and modulatory input specified correctly with respect to the underlying large scale neural model. The posterior parameter estimates for the coupling parameters are shown in black and grey; the values in brackets are the confidence that these values exceed a threshold of ln2/4 Hz. Coupling parameters exceeding threshold with a confidence of greater than 90% are shown in black. The posterior parameter estimates for the coupling parameters for direct visual inputs are shown next to the solid grey arrows. The posterior parameter estimates for the coupling parameters for modulatory effect of task (shapes vs. degraded stimuli) are shown next to the dotted grey markers. The values in brackets are the percentage confidence that these values are greater than zero. (Middle) Coupling parameters for the simple model, specified as a hierarchy. Inputs are specified and displayed as in the top panel. (Bottom) Coupling parameters for the simple model, specified with full inter area connectivity. Inputs are specified and displayed as in the top panel. Modified from Lee et al. (in press), which should be consulted for further details

actual anatomical connectivity (top); (2) a hierarchical arrangement (middle) and (3) an arrangement in which every region was reciprocally connected to every other (bottom). It can be seen that all the coupling parameters in the model on the top of Fig. 6 (the correctly specified model) exceeded threshold with greater than 90% confidence. In Fig 6 middle, the posterior probability of correctly specified connections exceeded 90% confidence, whereas the incorrect connections did not. The middle model shown in Fig. 6 reduces to the correct model when the coupling parameters that do not exceed threshold are excluded. Bayesian model comparison suggested that there was positive evidence for the correct model (the top model) over the two alternatives. For the

cases examined, Bayesian model comparison confirmed the validity of DCM in relation to our well established and comprehensive neuronal model.

It should also be mentioned that evaluation of interregional functional connectivity using EEG/MEG data is also performed by many groups, and computational neural modeling has been used to investigate these methods (e.g., David et al. 2004).

In summary, the use of large-scale modeling is starting to provide results that add support for some methods for computing functional and effective connectivity. However, the results so far also suggest that caution is needed in using fMRI-based calculations of functional and effective connectivity to infer the nature of interregional neural interactions from functional neuroimaging data.

4 Conclusions

In this chapter we have attempted to provide an overview of how computational modeling, especially those efforts employing large-scale neurobiologically realistic models, has in the last few years started to be used in conjunction with functional neuroimaging data. In our view, the major reason why neural modeling has become more central for interpreting functional brain imaging data is that there is a paradigm shift currently underway in cognitive neuroscience (see Fuster 2000), brought about primarily by the increased importance of functional neuroimaging studies. When the main sources of information about the neural basis of human cognitive operations came from the neuropsychological investigation of brain damaged patients and from electrophysiological and lesion studies in nonhuman preparations, scientific investigation focused on single brain regions and aimed at the notion of functional segregation (Zeki 1990). Functional brain imaging, especially PET and fMRI, demonstrated the importance of networks and has necessitated the development of network analysis methods. For the future, combined use of the different kinds of functional brain imaging methods – those, like fMRI, that provide good spatial information and those, like MEG, that provide good temporal information – will, in our view, necessitate even more intensive use of computational neural dynamic models.

Acknowledgements

This work was supported by the NIDCD Intramural Research Program.

References

Anderson JR, Albert MV, Fincham JM (2005) Tracing problem solving in real time: fMRI analysis of the subject-paced Tower of Hanoi. J Cogn Neurosci 17, 1261–1274

Anderson JR, Qin Y, Sohn MH, Stenger VA, Carter CS (2003) An information-processing model of the BOLD response in symbol manipulation tasks. Psychon Bull Rev 10, 241–261

Arbib MA, Bischoff A, Fagg AH, Grafton ST (1995) Synthetic PET: Analyzing large-scale properties of neural networks. Human Brain Mapp. 2, 225–233

Arbib MA (Ed.) (2003) The Handbook of Brain Theory and Neural Networks. MIT Press, Cambridge, MA.

Brown JW, Braver TS (2005) Learned predictions of error likelihood in the anterior cingulate cortex. Science 307, 1118–1121

Buechel C, Coull JT, Friston KJ (1999) The predictive value of changes in effective connectivity for human learning. Science 283, 1538–1541

Chadderdon GL, Sporns O (2006) A large-scale neurocomputational model of task-oriented behavior selection and working memory in the prefrontal cortex. J. Cogn. Neurosci. 18, 242–257

Ciocca V, Bregman AS (1987) Perceived continuity of gliding and steady-state tones through interrupting noise. Percept. Psychophys. 42, 476–484

Coltheart M, Rastle K, Perry C, Langdon R, Ziegler J (2001) DRC: A dual route cascaded model of visual word recognition and reading aloud. Psychol. Rev. 108, 204–256

Dannenbring GL (1976) Perceived auditory continuity with alternately rising and falling transitions. Can. J. Psychol. 30, 99–114

David O, Cosmelli D, Friston KJ (2004) Evaluation of different measures of functional connectivity using a neural mass model. Neuroimage 21, 659–673

David O, Friston KJ (2003) A neural mass model for MEG/EEG: coupling and neuronal dynamics. Neuroimage 20, 1743–1755

David O, Harrison L, Friston KJ (2005) Modelling event-related responses in the brain. Neuroimage 25, 756–770

Dayan P, Abbott LF (2001) Theoretical Neuroscience. MIT Press, Cambridge, MA.

Deco G, Rolls ET, Horwitz B (2004) 'What' and 'where' in visual working memory: a computational neurodynamical perspective for integrating fMRI and singe-cell data. J. Cogn. Neurosci. 16, 683–701

Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci 4, 2051–2062

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Ann. Rev. Neurosci. 18, 193–222

Desimone R, Ungerleider LG (1989). Neural mechanisms of visual processing in monkeys. In Goodglass H & Damasio AR (Eds.), Handbook of Neuropsychology 267–300. Elsevier, Amsterdam

Dominey PF, Arbib MA (1992) A cortico-subcortical model for generation of spatially accurate sequential saccades. Cereb. Cortex 2, 153–175

Frackowiak RSJ, Friston KJ, Frith CD, Dolan RJ, Price CJ, Zeki S, Ashburner J, Penny W (Eds.) (2004) Human Brain Function. Elsevier, Academic Press, San Diego, CA

Friston KJ (1994) Functional and effective connectivity in neuroimaging: a synthesis. Human Brain Mapp. 2, 56–78

Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ (1991) Investigating a network model of word generation with positron emission tomography. Proc. R. Soc. Lond. B 244, 101–106

Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. Neuroimage 19, 1273–1302

Funahashi S, Bruce C, Goldman-Rakic PS (1990) Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms. J. Neurophysiol. 63, 814–831

Fuster JM (2000) The module: Crisis of a paradigm (Review of The New Cognitive Neurosciences, 2nd Ed., edited by M. S. Gazzaniga). Neuron 26, 51–53

Gitelman DR, Penny WD, Ashburner J, Friston KJ (2003) Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. Neuroimage 19, 200–207

Griffiths TD, Warren JD (2004) What is an auditory object? Nat Rev Neurosci 5, 887–892

Harrison RV, Harel N, Hamrahi H, Panesar J, Mori N, Mount RJ (2000) Local haemodynamic changes associated with neural activity in auditory cortex. Acta Oto-laryngol. 120, 255–258

Haxby JV, Horwitz B, Ungerleider LG, Maisog JM, Pietrini P, Grady CL (1994) The functional organization of human extrastriate cortex: A PET-rCBF study of selective attention to faces and locations. J. Neurosci. 14, 6336–6353

Haxby JV, Ungerleider LG, Horwitz B, Rapoport SI, Grady CL (1995) Hemispheric differences in neural systems for face working memory: A PET-rCBF Study. Human Brain Mapp. 3, 68–82

Horwitz B (1990) Simulating functional interactions in the brain: A model for examining correlations between regional cerebral metabolic rates. Int. J. Biomed. Comput. 26, 149–170

Horwitz B (1994) Data analysis paradigms for metabolic-flow data: Combining neural modeling and functional neuroimaging. Human Brain Mapp. 2, 112–122

Horwitz B (2003) The elusive concept of brain connectivity. Neuroimage 19, 466–470

Horwitz B (2004) Relating fMRI and PET signals to neural activity by means of large-scale neural models. Neuroinformatics 2, 251–266

Horwitz B, Friston KJ, Taylor JG (2000) Neural modeling and functional brain imaging: an overview. Neural Networks 13, 829–846

Horwitz B, Glabus M (2005) Neural modeling and functional brain imaging: The interplay between data-fitting and simulation approaches. Int. Rev. Neurobiol. 66, 267–290

Horwitz B, Poeppel D (2002) How can EEG/MEG and fMRI/PET data be combined? Human Brain Mapp. 17, 1–3

Horwitz B, Soncrant TT, Haxby JV (1992). Covariance analysis of functional interactions in the brain using metabolic and blood flow data. In Gonzalez-Lima F, Finkenstaedt T, Scheich H (Eds.), Advances in Metabolic Mapping Techniques for Brain Imaging of Behavioral and Learning Functions 189–217. Kluwer Academic Publishers, Dordrecht, The Netherlands

Horwitz B, Sporns O (1994) Neural modeling and functional neuroimaging. Human Brain Mapp. 1, 269–283

Horwitz B, Tagamets M-A (1999) Predicting human functional maps with neural net modeling. Human Brain Mapp. 8, 137–142

Horwitz B, Tagamets M-A, McIntosh AR (1999) Neural modeling, functional brain imaging, and cognition. Trends Cogn. Sci. 3, 91–98

Horwitz B, Warner B, Fitzer J, Tagamets M-A, Husain FT, Long TW (2005) Investigating the neural basis for functional and effective connectivity: Application to fMRI. Phil. Trans. Roy. Soc. Lond. B 360, 1093–1108

Hubel DH, Wiesel TN (1977) Functional architecture of macaque momkey visual cortex. Proc. R. Soc. Lond. B Biol. Sci. 198, 1–59

Husain FT, Lozito TP, Ulloa A, Horwitz B (2005) Investigating the neural basis of the auditory continuity illusion. J. Cogn. Neurosci. 17, 1275–1292

Husain FT, Tagamets M-A, Fromm SJ, Braun AR, Horwitz B (2004) Relating neuronal dynamics for auditory object processing to neuroimaging activity. NeuroImage 21, 1701–1720

Jansen BH, Rit VG (1995) Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biol. Cybern. 73, 357–366

Jirsa VK, Haken H (1997) A derivation of a macroscopic field theory of the brain from the quasi-microscropic neural dynamics. Physica D 99, 503–526

Jueptner M, Weiller C (1995) Does measurement of regional cerebral blood flow reflect synaptic activity? – Implications for PET and fMRI. NeuroImage 2, 148–156

Just MA, Carpenter PA, Varma S (1999) Computational modeling of high-level cognition and brain function. Human Brain Mapp. 8, 128–136

Kaas JH, Hackett TA, Tramo MJ (1999) Auditory processing in primate cerebral cortex. Curr. Opinion Neurobiol. 9, 164–170

Kikuchi Y, Horwitz B, Mishkin M (2004) A patch of neurons in the monkey's rostral superior temporal gyrus are activated by conspecific calls. Abstracts, Soc. Neurosci. 30, Abstract #650.10

Kikuchi-Yorioka Y, Sawaguchi T (2000) Parallel visuospatial and audiospatial working memory processes in the monkey dorsolateral prefrontal cortex. Nat Neurosci 3, 1075–1076.

Kronhaus DM, Willshaw DJ (2006) The Cingulate as a Catalyst Region for Global Dysfunction: a Dynamical Modelling Paradigm. Cereb. Cortex 16, 1212–1224

Kubovy M, Van Valkenburg D (2001) Auditory and visual objects. Cognition 80, 97–126.

Lauritzen M (2001) Relationship of spikes, synaptic activity, and local changes of cerebral blood flow. J Cereb Blood Flow Metab 21, 1367–1383

Lee L, Friston KJ, Horwitz B (2006) Large-scale neural models and dynamic causal modelling. Neuroimage 30, 1243–1254

Logothetis NK (2003) MR imaging in the non-human primate: studies of function and of dynamic connectivity. Curr. Op. Neurobiol. 13, 630–642

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature 412, 150–157.

May P, Tiitinen H, Ilmoniemi RJ, Nyman G, Taylor JG, Naatanen R (1999) Frequency change detection in human auditory cortex. J. Comput. Neurosci. 6, 99–120

McClelland JL, Rumelhart DE (1986) Parallel Distributed Processing. MIT Press, Cambridge, MA.

McIntosh AR (2000) Towards a network theory of cognition. Neural Netw 13, 861–870

McIntosh AR, Gonzalez-Lima F (1991) Structural modeling of functional neural pathways mapped with 2-deoxyglucose: effect of acoustic startle habituation on the auditory system. Brain Res. 547, 295–302

McIntosh AR, Gonzalez-Lima F (1994) Structural equation modeling and its application to network analysis in functional brain imaging. Human Brain Mapp. 2, 2–22

McIntosh AR, Grady CL, Ungerleider LG, Haxby JV, Rapoport SI, Horwitz B (1994) Network analysis of cortical visual pathways mapped with PET. J. Neurosci. 14, 655–666

Mendelson JR, Cynader MS (1985) Sensitivity of cat primary auditory cortex (A1) neurons to the direction and rate of frequency modulation. Brain Res. 327, 331–335

Nunez P (1981) Electric Fields of the Brain. Oxford Univ. Press, New York.

Penny WD, Stephan KE, Mechelli A, Friston KJ (2004) Comparing dynamic causal models. Neuroimage 22, 1157–1172

Petrides M (1994) Frontal lobes and behaviour. Curr Opin Neurobiol 4, 207–211

Postle BR, D'Esposito M (1999) "What"-Then-Where" in visual working memory: an event-related fMRI study. J. Cogn. Neurosci. 11, 585–597

Rao SC, Rainer G, Miller EK (1997) Integration of what and where in the primate prefrontal cortex. Science 276, 821–824

Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci U S A 97, 11800–11806.

Robinson PA, Rennie CJ, Rowe DL, O'Connor SC, Gordon E (2005) Multiscale brain modelling. Phil. Trans. Roy. Soc. Lond. B 360, 1043–1050

Robinson PA, Rennie CJ, Wright JJ, Bahramali H, Gordon E, Rowe DL (2001) Prediction of electroencephalographic spectra from neurophysiology. Phys. Rev. E 63, 021903

Rolls ET, Deco G (2002) Computational Neuroscience of Vision. Oxford University Press, Oxford.

Rolls ET, Treves A (1998) Neural Networks and Brain Function. Oxford University Press, Oxford.

Stam CJ, Pijn JP, Suffczynski P, Lopes da Silva FH (1999) Dynamics of the human alpha rhythm: evidence for non-linearity? Clin. Neurophysiol. 110, 1801–1813

Sugita Y (1997) Neural correlates of auditory induction in the cat cortex. Neuroreport 8, 1155–1159

Tagamets M-A, Horwitz B (1998) Integrating electrophysiological and anatomical experimental data to create a large-scale model that simulates a delayed match-to-sample human brain imaging study. Cereb. Cortex 8, 310–320

Tagamets M-A, Horwitz B (2001) Interpreting PET and fMRI measures of functional neural activity: The effects of synaptic inhibition on cortical activation in human imaging studies. Brain Res. Bull. 54, 267–273

Tononi G, Sporns O, Edelman GM (1992) Reentry and the problem of integrating multiple cortical areas: Simulation of dynamic integration in the visual system. Cereb. Cortex 2, 310–335

Ungerleider LG, Mishkin M (1982). Two cortical visual systems. In Ingle DJ, Goodale MA, Mansfield RJW (Eds.), Analysis of Visual Behavior 549–586. MIT Press, Cambridge

Wilson FAW, O Scalaidhe SP, Goldman-Rakic PS (1993) Dissociation of object and spatial processing domains in primate prefrontal cortex. Science 260, 1955–1958

Wilson HR, Cowan JD (1972) Excitatory and inhibitory interactions in localized populations of model neurons. Biophys. J. 12, 1–24

Zeki S (1990) Functional specialization in the visual cortex: The generation of separate constructs and their multistage integration. In Edelman GM, Gall WE, Cowan WM (Eds.), Signal and Sense 85–130. Wiley-Liss, New York

Models of Effective Connectivity in Neural Systems

Klaas Enno Stephan and Karl J Friston

Wellcome Dept. of Imaging Neuroscience, Institute of Neurology, University College London, 12 Queen Square, WC1N 3BG, London, UK.

It is a longstanding scientific insight that understanding processes that result from the interaction of multiple elements require mathematical models of system dynamics (von Bertalanffy 1969). This notion is an increasingly important theme in neuroscience, particularly in neuroimaging, where causal mechanisms in neural systems are described in terms of effective connectivity. Here, we review established models of effective connectivity that are applied to data acquired with positron emission tomography (PET), functional magnetic resonance imaging (fMRI), electroencephalography (EEG) or magnetoencephalography (MEG). We start with an outline of general systems theory, a very general framework for formalizing the description of systems. This framework will guide the subsequent description of various establishd models of effective connectivity, including structural equation modeling (SEM), multivariate autoregressive modeling (MAR) and dynamic causal modeling (DCM). We focus particularly on DCM which distinguishes between neural state equations and a biophysical forward model that translates neural activity into a measured signal. After presenting some examples of applications of DCM to fMRI and EEG data, we conclude with some thoughts on pharmacological and clinical applications of models of effective connectivity.

1 General Systems Theory

The central goal of most scientific disciplines is to understand *systems*, i.e. ensembles of interacting elements. Today, this statement sounds almost trivial, yet in biology at least, the importance of the systems concept has been established only relatively recently. A key figure was Ludwig von Bertalanffy, a biologist and philosopher, who wrote a series of seminal articles in the first half of the 20^{th} century in which he argued that complex phenomena in biology (and indeed any other scientific field) invariably result from systems and could only be understood properly through a mathematical description of how system behavior emerged from the interactions of its constituent

elements. Demonstrating the existence of *system isomorphisms*, i.e. general mathematical descriptions that explained the dynamic behavior of very different kind of systems at different scales and across fields as diverse as physics, biology, economy and sociology, he introduced a very general framework that became known as *general system theory* (see the collection of essays in von Bertalanffy 1969). By the 1940s, the systems concept had experienced a scientific breakthrough in biology and led to the rise of *cybernetics*, "the science of control and communication in the animal and the machine" (Wiener 1948; Ashby 1956).

Today, biology uses the systems concept to address questions at all levels, from the molecular level to whole organisms and populations. The systems concept is now so omnipresent in biology that a recent special issue of the journal *Science* on systems biology renewed von Bertalanffy's (1969) previous diagnosis: "The [systems] concept has pervaded all fields of science and penetrated into popular thinking, jargon, and mass media" (Chong & Ray 2002).

But what exactly is a "system" and why is the systems concept so useful for framing scientific questions? A general, yet informal, definition is that a system is a set of elements which interact with each other in a spatially and temporally specific fashion. Before we attempt a formal definition of a system in the next section, let us remind ourselves that one of the classic scientific methods is to "analyze" a given phenomenon, i.e. to break it down into atomic units and processes that can be investigated independently of each other. This approach is appealing because it reduces a complex problem to a set of simpler problems, each of which can be addressed under conditions which can be controlled more easily for potentially confounding influences. For example, if one wanted to understand the physiological properties of a single neuron, one might decide to isolate it from its environment (e.g. let it grow in a dish) and then map its responses to currents injected into various parts of its dendritic tree. Unfortunately, this analytic approach cannot fully predict the neuron's behavior when it is part of a neural system, e.g. in the brain, and thus interacts with other neurons. When part of a system, the response of an individual neuron to a particular synaptic input (or injected current) u_1 depends on the spatial and temporal distribution of inputs $u_1 \ldots u_n$ that its dendritic tree receives from other neurons. If these additional inputs occur sufficiently close in time and space to u_1, they will affect the magnitude of the postsynaptic potential elicited by u_1, either linearly (by spatio-temporal summation) or nonlinearly (e.g. by changing the opening probability of voltage-gated channels) (Magee & Johnston 2005). In other words, the connectivity in the system mediates effects that cannot be predicted by studying a single neuron. Similar scenarios can be described for any other scientific field, for example biochemistry. Having studied a set of different biochemical processes in isolation, one would not necessarily be able to predict their collective dynamics. The problem is, as above, that different processes may interact, e.g. one process may change the substrate/product ratio of another process, or the efficacy of an enzyme that is relevant for a particular process may change due to the presence

of allosteric (in)activators that are produced by a second process or due to dynamic changes in gene expression mediated by a third process.

In summary, the general problem of analytical procedures in science is that they are blind to predicting the consequences arising from interactions between the elements in a system. Analytical procedures therefore need to be complemented with a theoretical framework that takes into account both the connectivity between the elements and external perturbations in order to achieve a mechanistic explanation of the dynamics of the system as a whole. This framework is provided by general system theory.

2 A General Form for System Models

Why is it useful at all to strive for formal mathematical definitions of systems? First, as described below, it allows one to pinpoint precisely what is meant by structure, function, and structure-function-relationships. Second, it allows one to predict system behavior for situations in which the system has not been observed before (see Bossel 1992 for an impressive collection of examples from biology). Third, it is the only way to fully understand how a system works and particularly, how system function could be restored if some of its components are rendered dysfunctional, e.g. by disease (Payne & Lomber 2001).

Here, we choose deterministic differential equations with time-invariant parameters as a mathematical framework; note that these are not the only possible mathematical representation of dynamic systems (see Bar-Yam 1997 for alternatives). The underlying concept, however, is quite universal: a *system* is defined by a set of elements with n time-variant properties altogether that interact with each other. Each time-variant property x_i $(1 \leq i \leq n)$ is called a *state variable*, and the n-vector $x(t)$ of all state variables in the system is called the *state vector* (or simply *state*) of the system at time t:

$$x(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \tag{1}$$

Taking an ensemble of interacting neurons as an example, the system elements would correspond to the individual neurons, each of which is represented by one or several state variables. These state variables could refer to various neurophysiological properties, e.g. postsynaptic potentials, status of ion channels, etc. This touches on an important distinction: in *system construction* (e.g. in engineering), the relevant state variables and their mutual dependencies are usually known; in *system identification* (e.g. when trying to understand a biological system), however, they are not known. This means that we always require a *model* of the system that represents our current hypothesis about the structure of the system and how its function emerges from that structure (the structure-function relationship, SFR).

The crucial point is that the state variables interact with each other, i.e. the evolution of each state variable depends on at least one other state variable. This mutual functional dependence between the state variables of the system is expressed in a very natural fashion by a set of ordinary differential equations that operate on the state vector:

$$\frac{dx}{dt} = \begin{bmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_n(x_1, \ldots, x_n) \end{bmatrix} = F(x) \tag{2}$$

However, this description is not yet sufficient. First of all, the specific form of the dependencies f_i needs to be specified, i.e. the nature of the causal relations between state variables. This requires a set of parameters θ which determine the form and strength of influences between state variables. In neural systems, these parameters usually correspond to time constants or strengths of the connections between the system elements. And second, in the case of non-autonomous systems (and these are the ones of interest to biology) we need to consider the input into the system, e.g. sensory information entering the brain. We represent the set of all m known inputs by the m-vector function $u(t)$. Extending (2) accordingly leads to a general state equation for non-autonomous deterministic systems

$$\frac{dx}{dt} = F(x, u, \theta) \tag{3}$$

where θ is the parameter vector of the system. Such a model provides a causal description of how system dynamics results from system structure, because it describes (i) when and where external inputs enter the system and (ii) how the state changes induced by these inputs evolve in time depending on the system's structure. As explained below in more detail in Sect. 3, (3) therefore provides a general form for models of *effective connectivity* in neural systems, i.e. the causal influences that neural units exert over another (Friston 1994).

We have made two main assumptions to simplify the exposition. First, it is assumed that all processes in the system are deterministic and occur instantaneously. Random components (noise) and delays could be accounted for by using stochastic differential equations and delay differential equations, respectively. Second, we assume that we know the inputs that enter the system. This is a tenable assumption in neuroimaging because the inputs are experimentally controlled variables, e.g. changes in stimuli or instructions.[1]

[1] Note that using time-invariant dependencies f_i and parameters θ is neither an assumption nor a restriction. Although the mathematical form of f_i per se is static, the use of time-varying inputs u allows for dynamic changes in what components of f_i are "activated". For example, using box-car functions that are multiplied with the different terms of a polynomial function one can induce changes from linear to nonlinear behavior (and vice versa) over time. Also, there is no principled distinction between states and time-invariant parameters. Therefore, estimating time-varying parameters can be treated as a state estimation problem.

On the basis of the general system description provided by (3) we can now state accurately, given a particular system model, what we mean by structure, function, and structure-function relationships (see Stephan 2004 for more details):

- S*tructure* is defined by the time-invariant components of the system, i.e. the binary nature of θ (which connections exist and which do not; see (8)) and the mathematical form of the state variable dependencies f_i.
- *Function* refers to those time-variant components of the system model that are conditional on its structure, i.e. $x(t)$, but not $u(t)$.
- The *structure-function relationship* (SFR) is represented by F: integrating F in time determines the temporal evolution of the system state x from time $t=0$ up to a time point τ, given an initial state $x(0)$:

$$x(\tau) = x(0) + \int_0^\tau F(x, u, \theta)dt \tag{4}$$

In other words, given a particular temporal sequence of inputs $u(t)$, (4) provides a complete description of how the dynamics of the system (i.e. the trajectory of its state vector x in time) results from its structure and initial state.

3 Functional Integration and Effective Connectivity are Assessed through System Models

Modern cognitive neuroscience has adopted an explicit system perspective. A commonly accepted view is that the brain regions that constitute a given system are computationally specialized, but that the exact nature of their individual computations depends on context, e.g. the inputs from other regions. The aggregate behavior of the system depends on this *neural context*, the context-dependent interactions between the system components (McIntosh 2000; see also the chapter by Bressler & McIntosh in this volume). An equivalent perspective is provided by the twin concepts of functional specialization and functional integration (Friston 2002). *Functional specialization* assumes a local specialization for certain aspects of information processing but allows for the possibility that this specialization is anatomically segregated across different cortical areas. The majority of current functional neuroimaging experiments have adopted this view and interpret the areas that are activated by a certain task component as the elements of a distributed system. However, this explanation is incomplete as long as no insight is provided into how the locally specialized computations are bound together by context-dependent interactions among these areas, i.e. the *functional integration* within the system. This functional integration within distributed neural systems can be characterized in two ways, functional connectivity and effective connectivity.

Functional connectivity has been defined as the temporal correlation between regional time series (Friston 1994). Analyses of functional connectivity therefore do not incorporate any knowledge or assumptions about the structure and the SFR of the system of interest. Depending on the amount of knowledge about the system under investigation, this can either be a strength or a weakness. If the system is largely unknown, functional connectivity approaches are useful because they can be used in an exploratory fashion, either by computing functional connectivity maps with reference to a particular seed region (Horwitz et al. 1998; McIntosh et al. 2003; Stephan et al. 2001a) or using a variety of multivariate techniques that find sets of voxels whose time series represent distinct (orthogonal or independent) components of the covariance structure of the data (Friston & Büchel 2004; McIntosh & Lobaugh 2004). The information from these analyses can then be used to generate hypotheses about the system. Conversely, given sufficient information about the system structure and a specific hypothesis about the SFR of the system, models of effective connectivity are more powerful. Here, we only deal with models of effective connectivity. For analyses of functional connectivity, please see the chapters by Salvador et al., Bressler & McIntosh and Sporns & Tononi in this volume.

Effective connectivity has been defined by various authors, but in complementary ways. A general definition is that effective connectivity describes the causal influences that neural units exert over another (Friston 1994). Other authors have proposed that "effective connectivity should be understood as the experiment- and time-dependent, simplest possible circuit diagram that would replicate the observed timing relationships between the recorded neurons" (Aertsen & Preißl 1991). Both definitions emphasize that determining effective connectivity requires a causal model of the interactions between the elements of the neural system of interest. Such a causal model has to take into account the external inputs that perturb the system and the anatomical connections by which neural units influence each other. In other words, any such model is a special case of the general system model as described in Sect. 2 and formalized by (3).

The equations presented in Sect. 2 are extremely general. To illustrate how the concept of effective connectivity emerges naturally from system models, we discuss the special case of a linear dynamic system. Although most natural phenomena are of a nonlinear nature, linear models play an important role in systems science because (i) they are analytically tractable, and (ii) given sufficiently long observation periods and non-negligible external input, their dynamics are largely independent of the initial state (Bossel 1992). Therefore nonlinear systems are usually investigated in restricted sub-spaces of interest, using linear models as local approximations. The following model of n inter-

acting brain regions is a simple linear case of (3) which uses a single state variable per region and m external inputs:

$$\begin{bmatrix} \frac{dx_1}{dt} \\ \vdots \\ \frac{dx_n}{dt} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} c_{11} & \cdots & c_{1m} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nm} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} \tag{5}$$

In this model the change of any given element depends on the state of the other system elements and on external inputs which affect it directly or indirectly. This system model can be written in compact matrix form as

$$F(x) = \frac{dx}{dt} = Ax + Cu \tag{6}$$

where the non-zero values of A and C represent the parameters of the system (i.e. θ in (3)) and the state of the system at time point τ can be obtained by integration (compare (4))

$$x(\tau) = e^{A\tau}x(0) + \int_0^\tau e^{A(\tau - t)}Cu(t)dt \tag{7}$$

where e^{At} is the matrix exponential (Bossel 1992). In this model, the system's behavior has two separable components: intrinsically sustained dynamics (parameterized by matrix A) and dynamics enforced by external inputs (parameterized by matrix C). The first term of (6) says that the change of the state variable x_i is a linear mixture of all state variables in the system, weighted by the parameters a_{ij}. By defining a particular parameter a_{ij} to be zero, we disallow for a direct effect of x_j on x_i (see Fig. 1 for an example). Conversely, any non-zero parameter a_{ij} represents a causal influence of the dynamics of x_j on that of x_i. The binarized parameter matrix $\tilde{A}$

$$\tilde{A} = \chi(A) = \begin{bmatrix} \chi(a_{11}) & \cdots & \chi(a_{1n}) \\ \vdots & \ddots & \vdots \\ \chi(a_{n1}) & \cdots & \chi(a_{nn}) \end{bmatrix},$$
$$\chi(a) = \begin{cases} 1 \text{ if } a \neq 0 \\ 0 \text{ if } a = 0 \end{cases} \tag{8}$$

represents the *structural connectivity* of the system model (see the chapter by Sporns & Tononi in this volume on how patterns of anatomical connections constrain effective connectivity and thus the dynamics of neural systems). The definition of the structural connectivity is usually guided by anatomical investigations in primates (Stephan et al. 2001b, Kötter 2004; see the chapter

by Paus in this volume for alternative approaches in humans). The values of A represent the influences of system elements over each other and thus correspond to the *effective connectivity* within the system. Finally, the values of the matrix C represent the magnitude of the direct effects that external (e.g. sensory) inputs have on system elements. By setting a particular parameter c_{ij} to be zero, we disallow for a direct effect of the external input u_j on x_i (see Fig. 1 for an example). A and C represent the system parameters (θ) that one needs to estimate when fitting this model to measured data. Simple linear models of this kind have found widespread application in various scientific disciplines (von Bertalanffy 1969). In Sect. 6, we will see that Dynamic Causal Modelling (DCM, Friston et al. 2003) extends the above formulation by bilinear terms that model context-dependencies of intrinsic connection strengths.

It should be noted that the framework outlined here is concerned with dynamic systems in continuous time and thus uses differential equations. The same basic ideas, however, can also be applied to dynamic systems in discrete time (using difference equations), as well as to "static" systems where the system is at equilibrium at each point of observation. The latter perspective, which is useful for regression-like equations, is used by classic system models for functional neuroimaging data, e.g. psycho-physiological interactions (PPI; Friston et al. 1997), structural equation modeling (SEM; McIntosh et al. 1994; Büchel & Friston 1997) or multivariate autoregressive models (MAR; Harrison et al. 2003; Göbel et al. 2003). These will be described in the following sections.

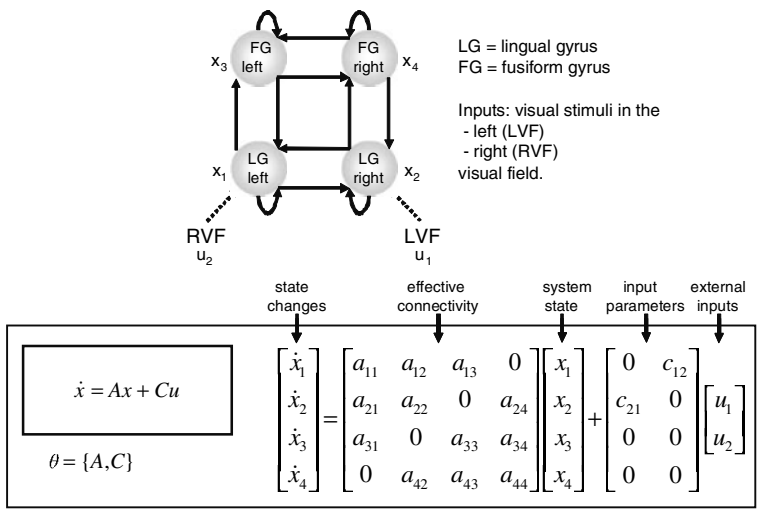


Fig. 1. A simple linear dynamic system as an example for a concrete implementation of (3), describing interactions between the lingual (LG) and the fusiform gyri (FG) in both hemispheres. The top panel shows the system structure and the sensory inputs (visual stimuli displayed in the left and right peripheral visual field) that perturb the system. The lower panel shows the state equation in matrix form

4 Psycho-Physiological Interactions (PPI)

PPI are one of the simplest models available to assess functional interactions in neuroimaging data (see Friston et al. 1997 for details). Given a chosen reference time series y_0 (obtained from a reference voxel or region), PPI computes whole-brain connectivity maps of this reference voxel with all other voxels y_i in the brain according to the regression-like equation

$$y_i = ay_0 + b(y_0 \times u) + cu + X\beta + e \tag{9}$$

Here, a is the strength of the intrinsic (context-independent) connectivity between y_0 and y_i. The bilinear term $y_0 \times u$ represents the interaction between physiological activity y_0 and a psychological variable u which can be construed as a contextual input into the system, modulating the connectivity between y_0 and y_i ($\times$ represents the Hadamard product, i.e. element-by element multiplication). The third term describes the strength c by which the input u determines activity in y_i directly, independent of y_0. Finally, β are parameters for effects of no interest X (e.g. confounds) and e is a Gaussian error term.

Notwithstanding the fact that this is a non-dynamic model, (9) contains the basic components of system descriptions as outlined in Sect. 2 and (3), and there is some similarity between its form and that of the state equation of DCM ((13), see below). However, since only pair-wise interactions are considered (i.e. separately between the reference voxel and all other brain voxels), this model is severely limited in its capacity to represent neural systems. This has also been highlighted in the initial description of PPIs (Friston et al. 1997). Although PPIs are not a proper system model, they have a useful role in exploring the functional interactions of a chosen region across the whole brain. This exploratory nature bears some similarity to analyses of functional connectivity. Unlike analyses of functional connectivity, however, PPIs model the contextual modulation of connectivity, and this modulation has a directional character, i.e. testing for a PPI from y_0 to y_i is not identical to testing for a PPI from y_i to y_0. This is because regressing $y_0 \times u$ on y_i is not equivalent to regressing $y_i \times u$ on y_0.

5 Structural Equation Modeling (SEM)

SEM has been an established statistical technique in the social sciences for several decades, but was only introduced to neuroimaging in the early 1990's by McIntosh & Gonzalez-Lima (1991). It is a multivariate, hypothesis-driven technique that is based on a structural model which represents the hypothesis about the causal relations between several variables (see McIntosh & Gonzalez-Lima 1994, Büchel & Friston 1997, Bullmore et al. 2000 and Penny et al. 2004a for methodological details). In the context of fMRI these variables are the measured BOLD (blood oxygen level dependent) time series

$y_1 \ldots y_n$ of n brain regions and the hypothetical causal relations are based on anatomically plausible connections between the regions. The strength of each connection $y_i \rightarrow y_j$ is specified by a so-called "path coefficient" which, by analogy to a partial regression coefficient, indicates how the variance of y_j depends on the variance of y_i if all other influences on y_j are held constant.

The statistical model of standard SEM implementations for neuroimaging data can be summarized by the equation

$$y = Ay + u \tag{10}$$

where y is a $n \times s$ matrix of n area-specific time series with s scans each, A is a $n \times n$ matrix of path coefficients (with zeros for non-existent connections), and u is a $n \times s$ matrix of zero mean Gaussian error terms, which are driving the modeled system ("innovations", see (11)). Note that the model on which SEM rests is a special case of the general equation for non-autonomous linear systems (with the exception that SEM is a static model and the inputs to the modeled system are random noise; compare (11) with (6)). Parameter estimation is achieved by minimizing the difference between the observed and the modeled covariance matrix $\mathcal{Y}$ of the areas (Bollen 1989). For any given set of parameters, Σ can be computed by transforming (10):

$$\begin{aligned} y &= (I - A)^{-1}u \\ \mathcal{Y} &= yy^T \\ &= (I - A)^{-1}uu^T(I - A)^{-1^T} \end{aligned} \tag{11}$$

where I is the identity matrix and T denotes the transpose operator. The first line of 11 can be understood as a generative model of how system function results from the system's connectional structure: the measured time series y results by applying a function of the inter-regional connectivity matrix, i.e. $(I - A)^{-1}$, to the Gaussian innovations u.

In the special case of fMRI, the path coefficients of a SEM (i.e. the parameters in A) describe the effective connectivity of the system across the entire experimental session. What one would often prefer to know, however, is how the coupling between certain regions changes as a function of experimentally controlled context, e.g. differences in coupling between two different tasks. Notably, SEM does not account for temporal order: if all regional time series were permuted in the same fashion, the estimated parameters would not change. In case of blocked designs, this makes it possible to proceed as if one were dealing with PET data, i.e. to partition the time series into condition-specific sub-series and fit separate SEMs to them. These SEMs can then be compared statistically to test for condition-specific differences in effective connectivity (for examples, see Büchel et al. 1999; Honey et al. 2002). An alternative approach is to augment the model with bilinear terms (cf. (9)) which represent the modulation of a given connection by experimentally controlled variables (e.g. Büchel & Friston 1997; Rowe et al. 2002). In this case, only a single SEM is fitted to the entire time series.

One limitation of SEM is that one is restricted to use structural models of relatively low complexity since models with reciprocal connections and loops often become non-identifiable (see Bollen 1989 for details). There are heuristics for dealing with complex models that use multiple fitting steps in which different parameters are held constant while changing others (see McIntosh et al. 1994 for an example).

6 Multivariate Autoregressive Models (MAR)

In contrast to SEM, autoregressive models explicitly address the temporal aspect of causality in time series. They take into account the causal dependence of the present on the past: each data point of a regional time series is explained as a linear combination of past data points from the same region. MAR models extend this approach to n brain regions, modeling the n-vector of regional signals at time t (y_t) as a linear combination of p past data vectors whose contributions are weighted by the parameter matrices A_i:

$$y_t = \sum_{i=1}^{p} y_{t-i} A_i + u_t \tag{12}$$

MAR models thus represent directed influences among a set of regions whose causal interactions are inferred via their mutual predictability from past time points. Although MAR is an established statistical technique, specific implementations for neuroimaging were suggested only relatively recently. Harrison et al. (2003) suggested a MAR implementation that allowed for the inclusion of bilinear variables representing modulatory effects of contextual variables on connections and used a Bayesian parameter estimation scheme specifically developed for MAR models (Penny & Roberts 2002). This Bayesian scheme also determined the optimal model order, i.e. the number of past time points (p in (12)) to be considered by the model. A complementary MAR approach, based on the idea of "Granger causality" (Granger 1969), was proposed by Goebel et al. (2003). In this framework, given two time-series y_1 and y_2, y_1 is considered to be caused by y_2 if its dynamics can be predicted better using past values from y_1 and y_2 as opposed to using past values of y_1 alone.

7 Dynamic Causal Modeling (DCM)

An important limitation of the models discussed so far is that they operate at the level of the measured signals. Taking the example of fMRI, the model parameters are fitted to BOLD series which result from a haemodynamic convolution of the underlying neural activity. Any inference about inter-regional connectivity obtained by PPI, SEM or MAR is only an indirect one because

these models do not include the forward model linking neuronal activity to the measured haemodynamic data. In the case of EEG, this forward model means there is a big difference between signals measured at each electrode and the underlying neuronal activity: changes in neural activity in different brain regions lead to changes in electric potentials that superimpose linearly. The scalp electrodes therefore record a mixture, with unknown weightings, of potentials generated by a number of different sources.

The causal architecture of the system that we would like to identify is expressed at the level of neuronal dynamics. Therefore, to enable inferences about connectivity between neural units we need models that combine two things: (i) a parsimonious but neurobiologically plausible model of neural population dynamics, and (ii) a biophysically plausible forward model that describes the transformation from neural activity to the measured signal. Such models make it possible to fit jointly the parameters of the neural and of the forward model such that the predicted time series are optimally similar to the observed time series. In principle, any of the models described above could be combined with a modality-specific forward model, and indeed, MAR models have previously been combined with linear forward models to explain EEG data (Yamashita et al. 2004). So far, however, Dynamic Causal Modeling (DCM) is the only approach where the marriage between models of neural dynamics and biophysical forward models is a mandatory component. DCM has been implemented both for fMRI (Friston et al. 2003) and EEG/MEG data (David et al. 2006; Kiebel et al. 2006). These modality-specific implementations are briefly summarized in the remainder of this section (see Fig. 2 for a conceptual overview).

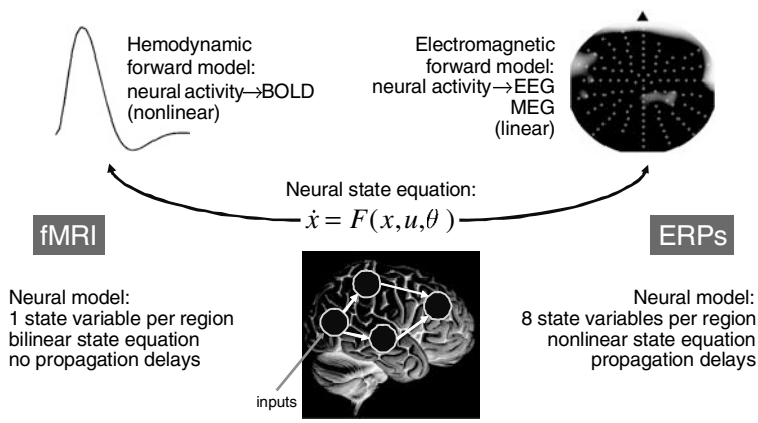


Fig. 2. A schematic overview that juxtaposes properties of DCM for fMRI and ERPs, respectively. It illustrates that DCM combines a model of neural population dynamics, following the generic form of (3), with a modality-specific biophysical forward model. Given appropriate formulations of the neural and the forward model, DCM can be applied to any kind of measurement modality

7.1 DCM for fMRI

DCM for fMRI uses a simple model of neural dynamics in a system of n interacting brain regions. It models the change of a neural state vector x in time, with each region in the system being represented by a single state variable, using the following bilinear differential equation:

$$\begin{aligned}\frac{dx}{dt} &= F(x, u, \theta^n) \\ &= \left(A + \sum_{j=1}^{m} u_j B^{(j)}\right) x + Cu \end{aligned} \tag{13}$$

Note that this neural state equation follows the general form for deterministic system models introduced by (3), i.e. the modeled state changes are a function of the system state itself, the inputs u and some parameters θ^n that define the functional architecture and interactions among brain regions at a neuronal level (n in θ^n is not an exponent but a superscript that denotes "neural"). The neural state variables represent a summary index of neural population dynamics in the respective regions. The neural dynamics are driven by experimentally controlled external inputs that can enter the model in two different ways: they can elicit responses through direct influences on specific regions (e.g. evoked responses in early sensory cortices; the C matrix) or they can modulate the coupling among regions (e.g. during learning or attention; the B matrices).

Equation (13) is a bilinear extension of (6) that was introduced earlier as an example of linear dynamic systems. Given this bilinear form, the neural parameters $\theta^n = \{A, B, C\}$ can be expressed as partial derivatives of F:

$$\begin{aligned} A &= \left.\frac{\partial F}{\partial x}\right|_{u=0} \\ B^{(j)} &= \frac{\partial^2 F}{\partial x \partial u_j} \\ C &= \left.\frac{\partial F}{\partial u}\right|_{x=0} \end{aligned} \tag{14}$$

The matrix A represents the effective connectivity among the regions in the absence of input, the matrices $B^{(j)}$ encode the change in effective connectivity induced by the j^{th} input u_j, and C embodies the strength of direct influences of inputs on neuronal activity (see Fig. 3 for a concrete example and compare it to Fig. 1).

DCM for fMRI combines this model of neural dynamics with an experimentally validated haemodynamic model that describes the transformation of neuronal activity into a BOLD response. This so-called "Balloon model" was initially formulated by Buxton et al. (1998) and later extended by

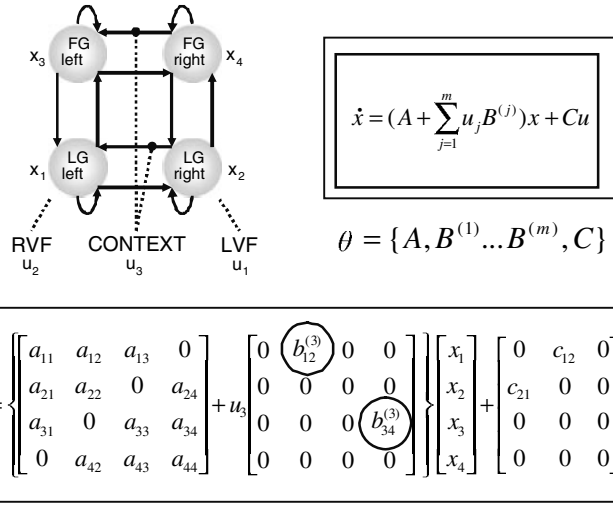


Fig. 3. A simple bilinear extension of the linear dynamic system shown in Fig. 1. This is an example for a concrete implementation of the neural state equation of DCM for fMRI. Note the role of the bilinear terms which model context-dependent (additive) changes of the strengths of the connections from the right to the left hemisphere (circled elements in the B matrix)

Friston et al. (2000). Briefly, it consists of a set of differential equations that describe the relations between four haemodynamic state variables, using five parameters (θ^h). More specifically, changes in neural activity elicit a vasodilatory signal that leads to increases in blood flow and subsequently to changes in blood volume and deoxyhemoglobin content. The predicted BOLD signal is a non-linear function of blood volume and deoxyhemoglobine content. Details of the haemodynamic model can be found in other publications (Friston et al. 2000; Stephan et al. 2004). Figure 4 provides a conceptual overview of DCM for fMRI.

The combined neural and haemodynamic parameter set $\theta = \{\theta^n, \theta^h\}$ is estimated from the measured BOLD data, using a fully Bayesian approach with empirical priors for the haemodynamic parameters and conservative shrinkage priors for the coupling parameters. Details of the parameter estimation scheme, which rests on a gradient ascent procedure embedded into an expectation maximization (EM) algorithm and uses a Laplace (i.e. Gaussian) approximation to the true posterior, can be found in Friston (2002). Eventually, the posterior distributions of the parameter estimates can be used to test hypotheses about connection strengths. Usually, these hypotheses concern context-dependent changes in coupling. If there is uncertainty about the connectional structure of the modeled system, or if one would like to compare competing hypotheses (represented by different DCMs), a Bayesian model selection procedure can be used to find the DCM that exhibits an optimal balance between model fit and model complexity (Penny et al. 2004b).

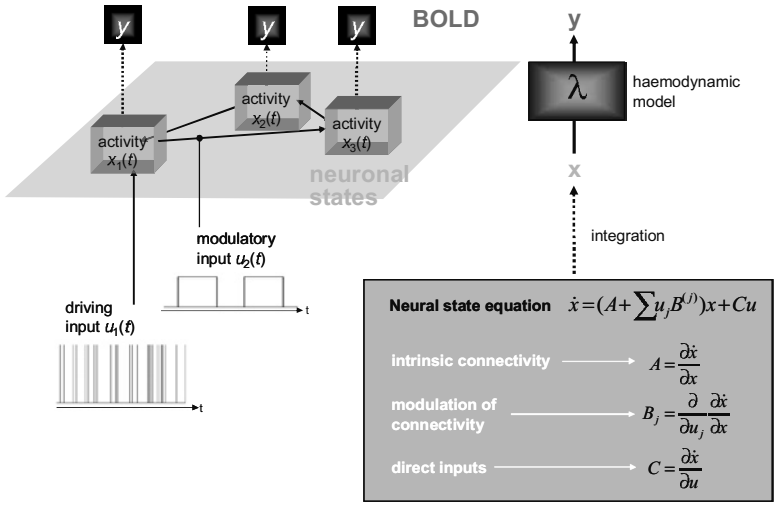


Fig. 4. Schematic summary of DCM for fMRI. The dynamics in a system of interacting neuronal populations (left panel), which are not directly observable by fMRI, are modeled using a bilinear state equation (right panel). Integrating the state equation gives predicted neural dynamics (x) which are transformed into predicted BOLD responses (y) by means of a haemodynamic forward model (λ). Neural and haemodynamic parameters are adjusted jointly such that the differences between predicted and measured BOLD series are minimized. The neural dynamics are determined by experimental manipulations that enter the model in the form of external inputs. Driving inputs (u_1; e.g. sensory stimuli) elicit local responses which are propagated through the system according to the intrinsic connections. The strengths of these connections can be changed by modulatory inputs (u_2; e.g. changes in task, attention, or due to learning). Note that in this figure the structure of the system and the scaling of the inputs have been chosen arbitrarily

7.2 DCM for Event-Related Potentials (ERPs)

ERPs as measured with EEG or MEG have been used for decades to study electrophysiological correlates of cognitive operations. Nevertheless, the neurobiological mechanisms that underlie their generation are still largely unknown. DCM for ERPs was developed as a biologically plausible model to understand how event-related responses result from the dynamics in coupled neural ensembles (David et al. 2006).

DCM for ERPs rests on a neural mass model, developed by David & Friston (2003) as an extension of the model by Jansen & Rit (1995), which uses established connectivity rules in hierarchical sensory systems (Felleman & Van Essen 1992) to assemble a network of coupled cortical sources. These rules characterize connections with respect to their laminar patterns of origin and termination and distinguish between (i) forward (or bottom-up) connections originating in agranular layers and terminating in layer 4, (ii) backward (or top-down) connections originating and terminating in agranular layers,

and (iii) lateral connections originating in agranular layers and targeting all layers. These long-range (extrinsic or inter-areal) cortico-cortical connections are excitatory, using glutamate as neurotransmitter, and arise from pyramidal cells.

Each region or source is modeled as a microcircuit following the model by David & Friston (2003). Three neuronal subpopulations are combined in this circuit and assigned to granular and supra-/infragranular layers. A population of excitatory pyramidal (output) cells receives inputs from inhibitory and excitatory populations of interneurons via intrinsic (intra-areal) connections. Within this model, excitatory interneurons can be regarded as spiny stellate cells found predominantly in layer 4 and in receipt of forward connections. Excitatory pyramidal cells and inhibitory interneurons are considered to occupy infra- and supragranular layers and receive backward and lateral inputs (see Fig. 5).

The neural state equations are summarized in Fig. 5. To perturb the system and model event-related responses, the network receives inputs via input connections. These connections are exactly the same as forward connections and deliver input u to the spiny stellate cells in layer 4. Input u represents afferent activity relayed by subcortical structures and are modelled as two parameterized components, a gamma density function (representing an event-related burst of input that is delayed and dispersed by subcortical synapses

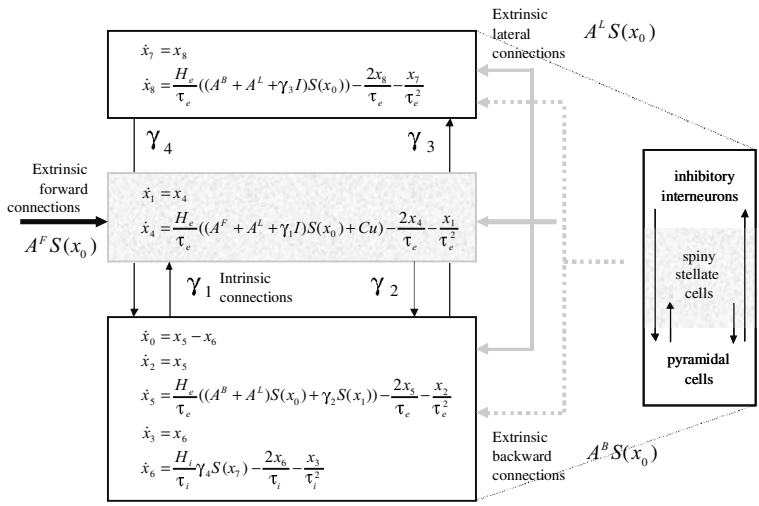


Fig. 5. Schematic of the neural model in DCM for ERPs. This schema shows the state equations describing the dynamics of a microcircuit representing an individual region (source). Each region contains three subpopulations (pyramidal, spiny stellate and inhibitory interneurons) that are linked by intrinsic connections and have been assigned to supragranular, granular and infragranular cortical layers. Different regions are coupled through extrinsic (long-range) excitatory connections that follow the laminar patterns of forward, backward and lateral connections, respectively

and axonal conduction) and a discrete cosine set (representing fluctuations in input over peristimulus time). The influence of this input on each source is controlled by a parameter vector C (see David et al. 2006 for details). Overall, the DCM is specified in terms of the state equations shown in Fig. 5 and a linear output equation

$$\begin{aligned} \frac{dx}{dt} &= f(x, u, \theta) \\ y &= Lx_0 + \varepsilon \end{aligned} \tag{15}$$

where x_0 represents the transmembrane potential of pyramidal cells and L is a lead field matrix coupling electrical sources to the EEG channels (Kiebel et al. 2006). In comparison to DCM for fMRI, the forward model is a simple linearity as opposed to the nonlinear haemodynamic model in DCM for fMRI. In contrast, the state equations of DCM for ERPs are much more complex and realistic (cf. Fig. 5). As an example, the state equation for the inhibitory subpopulation is

$$\begin{aligned} \frac{dx_7}{dt} &= x_8 \\ \frac{dx_8}{dt} &= \frac{H_e}{\tau_e}((A^B + A^L + \gamma_3 I)S(x_0)) - \frac{2x_8}{\tau_e} - \frac{x_7}{\tau_e^2} \end{aligned} \tag{16}$$

The parameter matrices A^F, A^B, A^L encode forward, backward and lateral connections respectively. Within each subpopulation, the dynamics of neural states are determined by two operators. The first transforms the average density of presynaptic inputs into the average postsynaptic membrane potential. This is modeled by a linear transformation with excitatory (e) and inhibitory (i) kernels parameterized by $H_{e,i}$ and $\tau_{e,i}$. $H_{e,i}$ control the maximum postsynaptic potential and $\tau_{e,i}$ represent lumped rate constants (i.e. lumped across dendritic spines and the dendritic tree). The second operator S transforms the average potential of each subpopulation into an average firing rate. This is assumed to be instantaneous and is a sigmoid function. Intra-areal interactions among the subpopulations depend on constants $\gamma_{1\ldots4}$ which control the strength of intrinsic connections and reflect the total number of synapses expressed by each subpopulation. In (16), the top line expresses the rate of change of voltage as a function of current. The second line specifies how current changes as a function of voltage, current and presynaptic input from extrinsic and intrinsic sources. For simplification, our description here has omitted the fact that in DCM for ERPs all intra- and inter-areal connections have conduction delays. This requires the use of delay differential equations (see David et al. 2006 for details).

For estimating the parameters from empirical data, a fully Bayesian approach is used that is analogous to that used in DCM for fMRI and is described in detail by David et al. (2006). The posterior distributions of the parameter estimates can be used to test hypotheses about the modeled

processes, particularly differences in inter-areal connection strengths between different trial types. As in DCM for fMRI, Bayesian model selection can be used to optimize model structure or compare competing scientific hypotheses (Penny et al. 2004b).

8 Application of System Models in Functional Neuroimaging: Present and Future

Models of functional integration, which were originally developed for electrophysiological data from multi-unit recordings (Gerstein and Perkel 1968), are now taking an increasingly prominent role in functional neuroimaging. This is because the emphasis of the scientific questions in cognitive neuroscience is shifting from *where* particular processes are happening in the brain to *how* these processes are implemented. With increasing use, a word of caution may be appropriate here: Models of effective connectivity are not very useful without precise a priori hypotheses about specific mechanisms expressed at the level of inter-regional coupling. Simply describing *patterns* of connectivity that require post hoc interpretation does not lead to a mechanistic understanding of the system of interest. What is needed are parsimonious, well-motivated models that test precise hypotheses about *mechanisms*, either in terms of changes in particular connection strengths as a function of experimental condition, time (learning) or drug, or in terms of comparing alternative explanations by model selection (for examples, see Büchel & Friston 1997; Büchel et al. 1999; Honey et al. 2003; McIntosh et al. 1994, 1998; Rowe et al. 2002; Stephan et al. 2003, 2005; Toni et al. 2002). Figure 6 shows an example of such a model (Friston et al. 2003) where the parameters are mechanistically meaningful.

This search for mechanisms seems particularly promising for pharmacological questions. Since many drugs used in psychiatry and neurology change synaptic transmission and thus functional coupling between neurons, their therapeutic effects cannot be fully understood without models of drug-induced connectivity changes in particular neural systems. So far, only relatively few studies have studied pharmacologically induced changes in connectivity, ranging from simple analyses of functional connectivity (e.g. Stephan et al. 2001a) to proper system models (e.g. Honey et al. 2003). As highlighted in a recent review by Honey and Bullmore (2004), an exciting possibility for the future is to use system models at the early stage of drug development to screen for substances that induce desired changes of connectivity in neural systems of interest with a reasonably well understood physiology. The success of this approach will partially depend on developing models that include additional levels of biological detail (e.g. effects of different neurotransmitters and receptor types) while being parsimonious enough to ensure mathematical identifiability and physiological interpretability; see Breakspear et al. (2003),

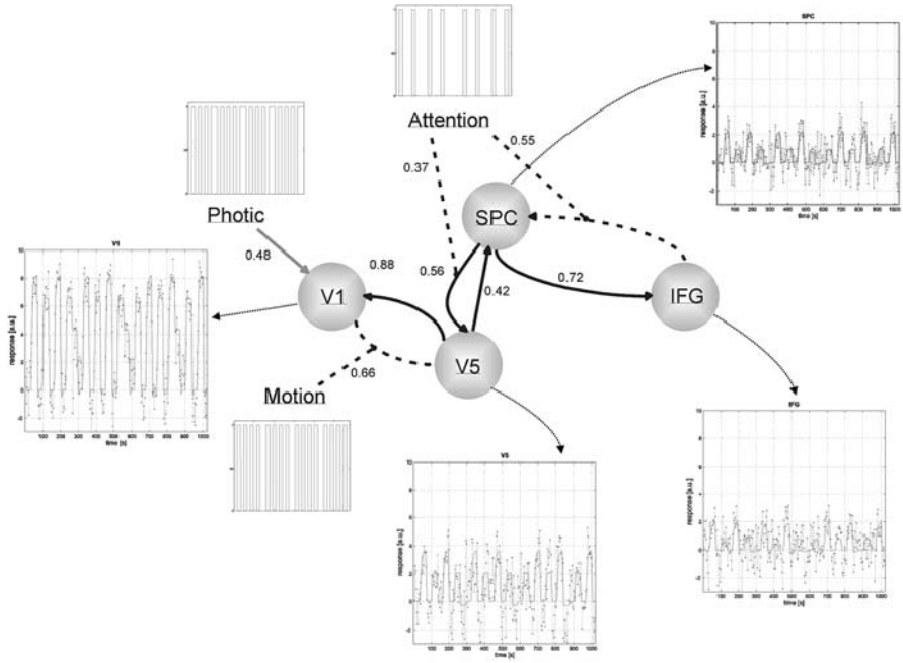


Fig. 6. DCM analysis of a single subject fMRI data from a study of attention to visual motion in which subjects viewed identical stimuli (radially moving dots) under different levels of attention to the stimuli (Büchel & Friston 1997). The model was introduced and described in detail by Friston et al. (2003). The figure is reproduced (with permission from Elsevier Ltd.) from Stephan et al. (2004). Only those conditional estimates are shown alongside their connections for which there was at least 90% confidence that they corresponded to neural transients with a half life shorter than 4 seconds. The temporal structure of the inputs is shown by box-car plots. Dashed arrows connecting regions represent significant bilinear affects in the absence of a significant intrinsic coupling. Fitted responses based upon the conditional estimates and the adjusted data are shown in the panels connected to the areas by dotted lines. The important parameters here are the bilinear ones. Note that while the intrinsic connectivity between areas V1 and V5 is non-significant and basically zero, motion stimuli drastically increase the strength of this connection, "gating" V1 input to V5. Top-down effects of attention are represented by the modulation of backward connections from the inferior frontal gyrus (IFG) to the superior parietal cortex (SPC) and from SPC to V5. See Penny et al. (2004b) and Stephan (2004) for a discussion how different neurophysiological mechanisms can be modeled with DCM

Harrison et al. (2005), Jirsa (2004) and Robinson et al. (2001) for examples that move in this direction.

Another important goal is to explore the utility of models of effective connectivity as diagnostic tools (Stephan 2004). This seems particularly attractive for psychiatric diseases whose phenotypes are often very heterogeneous and

where a lack of focal brain pathologies points to abnormal connectivity (dysconnectivity) as the cause of the illness. Given a pathophysiological theory of a specific disease, connectivity models might allow one to define an *endophenotype* of that disease, i.e. a biological marker at intermediate levels between genome and behaviour, which enables a more precise and physiologically motivated categorization of patients (Gottesman & Gould 2003). Such an approach has received particular attention in the field of schizophrenia research where a recent focus has been on abnormal synaptic plasticity leading to dysconnectivity in neural systems concerned with emotional and perceptual learning (Friston 1998; Stephan et al. 2006). A major challenge will be to establish neural systems models which are sensitive enough that their connectivity parameters can be used reliably for diagnostic classification and treatment response prediction of individual patients. Ideally, such models should be used in conjunction with paradigms that are minimally dependent on patient compliance and are not confounded by factors like attention or performance. Given established validity and sufficient sensitivity and specificity of such a model, one could use it in analogy to biochemical tests in internal medicine, i.e. to compare a particular model parameter (or combinations thereof) against a reference distribution derived from a healthy population (Stephan et al. 2006). Such procedures could help to decompose current psychiatric entities like schizophrenia into more well-defined subgroups characterized by common pathophysiological mechanisms and may facilitate the search for genetic underpinnings.

Acknowledgements

The authors are supported by the Wellcome Trust. We would like to thank our colleagues for stimulating discussions on modeling, particularly Lee Harrison, Will Penny, Olivier David, Stefan Kiebel, Randy McIntosh and Michael Breakspear.

References

Aertsen A, Preißl H (1991) Dynamics of activity and connectivity in physiological neuronal Networks. In: Schuster HG (ed.) Non Linear Dynamics and Neuronal Networks, VCH Publishers, New York, pp. 281–302

Ashby WR (1956) An introduction to cybernetics. Chapman & Hall, London

Bar-Yam Y (1997) Dynamics of complex systems. Addison-Wesley, Reading

Bollen KA (1989) Structural equations with latent variables. John Wiley, New York

Bossel H (1992) Modellbildung und Simulation. Vieweg, Braunschweig

Breakspear M, Terry JR, Friston KJ (2003) Modulation of excitatory synaptic coupling facilitates synchronization and complex dynamics in a biophysical model of neuronal dynamics. Network: Computation in Neural Systems 14, 703–732

Büchel C, Friston KJ (1997) Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cerebral Cortex 7, 768–778

Büchel C, Coull JT, Friston KJ (1999) The Predictive Value of Changes in Effective Connectivity for Human Learning. Science 283, 1538–1541

Bullmore ET, Horwitz B, Honey G, Brammer M, Williams S, Sharma T (2000) How good is good enough in path analysis of fMRI data? NeuroImage 11, 289–301

Buxton RB, Wong EC, Frank LR (1998) Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magnetic Resonance Medicine 39, 855–864

Chong L, Ray LB (2002) Whole-istic biology. Science 295, 1661

David O, Friston KJ (2003) A neural mass model for MEG/EEG: coupling and neuronal dynamics. NeuroImage 20, 1743–1755

David O, Kiebel SJ, Harrison LM, Mattout J, Kilner JM, Friston KJ (2006) Dynamic causal modeling of evoked responses in EEG and MEG. NeuroImage 30, 1255–1272

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cerebral Cortex 1, 1–47

Friston KJ (1994) Functional and effective connectivity in neuroimaging: a synthesis. Human Brain Mapping 2, 56–78

Friston KJ, Büchel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. NeuroImage 6, 218–229

Friston KJ (1998) The disconnection hypothesis. Schizophrenia Research 30, 115–125

Friston KJ, Mechelli A, Turner R, Price CJ (2000) Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. NeuroImage 12, 466–477

Friston KJ (2002) Beyond phrenology: What can neuroimaging tell us abut distributed circuitry? Annual Reviews in Neuroscience 25, 221–250

Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. NeuroImage 19, 1273–1302

Friston KJ, Büchel C (2004) Functional connectivity: eigenimages and multivariate analyses. In: Frackowiack R et al. (ed.) Human Brain Function, 2nd edition, Elsevier, New York, pp. 999–1018

Gerstein GL, Perkel DH (1969) Simultaneously recorded trains of action potentials: Analysis and functional interpretation. Science 164, 828–830

Goebel R, Roebroeck A, Kim DS, Formisano E (2003) Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. Magnetic Resonance Imaging 21, 1251–1261

Gottesman II, Gould TD (2003) The endophenotype concept in psychiatry: etymology and strategic intentions. American Journal of Psychiatry 160, 636–645

Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438

Harrison LM, Penny W, Friston KJ (2003) Multivariate autoregressive modeling of fMRI time series. NeuroImage 19, 1477–1491

Harrison LM, David O, Friston KJ (2005) Stochastic models of neuronal dynamics. Philosophical Transactions of the Royal Society London B Biological Sciences 360, 1075–1091

Honey GD, Fu CHY, Kim J, Brammer MJ, Croudace TJ, Suckling J, Pich EM, Williams SCR, Bullmore ET (2002) Effects of verbal working memory load on corticocortical connectivity modeled by path analysis of functional magnetic resonance imaging data. NeuroImage 17, 573–582

Honey GD, Suckling J, Zelaya F, Long C, Routledge C, Jackson S, Ng V, Fletcher PC, Williams SCR, BrownJ, Bullmore ET (2003) Dopaminergic drug effects on physiological connectivity in a human cortico-striato-thalamic system. Brain 126, 1767–1281

Honey G, Bullmore E (2004) Human pharmacological MRI. Trends in Pharmacological Sciences 25, 366–374

Horwitz B, Rumsey JM & Donohue BC (1998) Functional connectivity of the angular gyrus in normal reading and dyslexia. Proceedings of the National Academy of Sciences USA 95, 8939–8944

Jirsa VK (2004) Connectivity and dynamics of neural information processing. Neuroinformatics 2, 183–204

Kiebel SJ, David O, Friston KJ (2006) Dynamic causal modelling of evoked responses in EEG/MEG with lead-field parameterization. NeuroImage 30, 1273–1284

Kötter R (2004) Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. Neuroinformatics 2, 127–144

Magee JC, Johnston D (2005) Plasticity of dendritic function. Current Opinion in Neurobiology 15, 334–342

McIntosh AR, Gonzalez-Lima F (1991) Metabolic activation of the rat visual system by patterned light and footshock. Brain Research 547, 295–302

McIntosh AR, Gonzalez-Lima F (1994) Structural equation modeling and its application to network analysis in functional brain imaging. Human Brain Mapping 2, 2–22

McIntosh AR, Grady CL, Ungerleider LG, Haxby JV, Rapoport SI, Horwitz B (1994) Network analysis of cortical visual pathways mapped with PET. Journal of Neuroscience 14, 655–666

McIntosh AR, Cabeza RE, Lobaugh NJ (1998) Analysis of neural interactions explains the activation of occipital cortex by an auditory stimulus. Journal of Neurophysiology 80, 2790–2796

McIntosh AR (2000) Towards a network theory of cognition. Neural Networks 13, 861–870

McIntosh AR, Rajah MN, Lobaugh NJ (2003) Functional connectivity of the medial temporal lobe relates to learning and awareness. Journal of Neuroscience 23, 6520–6528

McIntosh AR, Lobaugh NJ (2004) Partial least squares analysis of neuroimaging data: applications and advances. NeuroImage 23, S250–S263

Payne BR, Lomber SG (2001) Reconstructing functional systems after lesions of cerebral cortex. Nature Reviews Neuroscience 2, 911–919

Penny WD, Roberts SJ (2002) Bayesian multivariate autoregressive models with structured priors. IEE Proceedings of Vision and Image Signal Processing 149, 33–41

Penny WD, Stephan KE, Mechelli A, Friston KJ (2004a) Modeling functional integration: a comparison of structural equation and dynamic causal models. NeuroImage 23, S264–274

Penny WD, Stephan KE, Mechelli A, Friston KJ (2004b) Comparing dynamic causal models. NeuroImage 22, 1157–1172

Robinson PA, Rennie CJ, Wright JJ, Bahramali H, Gordon E, Rowe DL (2001) Prediction of electroencephalographic spectra from neurophysiology. Physical Reviews E 63, 021903

Rowe JB, Stephan KE, Friston KJ, Frackowiak RJ, Lees A, Passingham RE (2002) Attention to action in Parkinson's disease. Impaired effective connectivity among frontal cortical regions. Brain 125, 276–289

Stephan KE, Magnotta VA, White TJ, Arndt S, Flaum M, O'Leary DS & Andreasen, NC (2001a) Effects of Olanzapine on cerebellar functional connectivity in schizophrenia measured by fMRI during a simple motor task. Psychological Medicine 31, 1065–1078

Stephan KE, Kamper L, Bozkurt A, Burns GAPC, Young MP, Kötter R (2001b) Advanced database methodology for the Collation of Connectivity data on the Macaque brain (CoCoMac). Philosophical Transactions of the Royal Society London B Biological Sciences 356, 1159–1186

Stephan KE, Marshall JC, Friston KJ, Rowe JB, Ritzl A, Zilles K, Fink GR (2003) Lateralized cognitive processes and lateralized task control in the human brain. Science 301, 384–386

Stephan KE (2004) On the role of general system theory for functional neuroimaging. Journal of Anatomy 205, 443–470

Stephan KE, Harrison LM, Penny WD, Friston KJ (2004) Biophysical models of fMRI responses. Current Opinion in Neurobiology 14, 629–635

Stephan KE, Penny WD, Marshall JC, Fink GR, Friston KJ (2005) Investigating the functional role of callosal connections with dynamic causal models. Annals of the New York Academy of Sciences 1064:16–36

Stephan KE, Baldeweg T, Friston KJ (2006) Synaptic plasticity and dysconnection in schizophrenia. Biological Psychiatry 59, 929–939

Toni I, Rowe JB, Stephan KE & Passingham RE (2002) Changes of cortico-striatal effective connectivity during visuo-motor learning. Cerebral Cortex 12: 1040–1047

von Bertalanffy L (1969) General System Theory. George Braziller, New York

Wiener N (1948) Cybernetics. Wiley, New York

Yamashita O, Galka A, Ozaki T, Biscay R, Valdes-Sosa P (2004) Recursive penalized least squares solution for dynamical inverse problems of EEG generation. Human Brain Mapping 21, 221–235

Multichannel Data Analysis in Biomedical Research

Maciej Kamiński

Department of Biomedical Physics, Institute of Experimental Physics, Warsaw University, ul. Hoża 69, 00-681, Warszawa, Poland

1 Introduction

Multivariate data can be encountered in many fields of science or engineering. Any experiment or measurement with several quantities simultaneously recorded delivers multivariate datasets. This is especially true in biomedical investigations where most of the recordings are nowadays multichannel. EEG equipment is able to record signals from still more and more electrodes, but the notion is not confined only to electrode recordings: even fMRI data can be treated as a multivariate set of voxels changing their state in time. Such multichannel recordings are intended to deliver more information about the investigated object. However, the amount of knowledge obtained from the data analysis depends on the analysis method used. An analytical tool which is improperly applied may not give the correct answers, in fact, it may deliver incomplete or false information unbeknownst to the researcher.

The aim of this chapter is to describe the main aspects that are important in processing multivariate data. In the first, theoretical part, the specific properties of that type of data will be described and related functions will be introduced. In the second part, examples of treating typical problems arising during the analysis will be presented from a practical point of view and solutions will be proposed. The linear modeling approach will be discussed. This chapter can be viewed as a guide explaining basic properties of multivariate datasets and problems specific for multivariate analysis presented with selected relevant examples. For the issues considered in the text references to the literature for further, more comprehensive reading will be given.

2 Terminology

Let $\mathbf{X}$ denote a multivariate stochastic process containing k subprocesses. A value of the process at a time t can be expressed as ($^{\mathrm{T}}$ denotes transposition)

$$\mathbf{X}(t) = (X_1(t), X_2(t), \ldots, X_k(t))^{\mathrm{T}} \tag{1}$$

In the further discussion we assume that every process X_i has a zero mean. In the data collecting practice, data acquisition requires sampling values of each process X_i at certain equally spaced time points. This results in a time domain representation of multiple time series with a certain sampling period Δt. The quantity $f_s = 1/\Delta t$ is the sampling frequency. Although many characteristics of the measured signals can be evaluated directly from the data in the time domain, quite often spectral properties are of primary interest. The signal $\mathbf{X}(t)$ can be transformed into the frequency domain by application of the Fourier transform. The frequency representation of the signal, $\tilde{\mathbf{X}}(f)$, is a complex valued function describing amplitudes and phases of frequency components of the signal at the frequency f. Later in the text we will omit the tilde above symbols representing frequency representations of respective time domain quantities, remembering that both symbols (like $\mathbf{X}(t)$ and $\mathbf{X}(f)$) signify different quantities. The power spectral density matrix of the signal $\mathbf{X}$ is defined as

$$\mathbf{S}(f) = \mathbf{X}(f)\mathbf{X}^*(f), \tag{2}$$

where the superscript asterisk represents transposition and complex conjugate. The matrix $\mathbf{S}$ is often simply called a spectrum. Its diagonal elements are called the auto-spectra, the off-diagonal elements are called the cross-spectra.

3 Multivariate Analysis

Typically, when analyzing a univariate data record (a single time series), several quantities describing properties of the data are estimated. In the case of a multivariate dataset the same estimations can be repeated for each data channel separately. Information gathered from each channel can be very useful, for instance in mapping procedures. However, a multivariate dataset contains additional information about relations between channels. To evaluate these so called cross-relations, specific functions depending on two (or more) signals simultaneously were defined.

Covariance (or correlation in a normalized version) is a function describing common trends in behavior of two time series X_i and X_j. The function operates in the time domain. The simplest general formula can be written as:

$$cov(X_i, X_j, s) = \frac{1}{\tau} \sum_m X_i(m) X_j(m+s) \tag{3}$$

Depending on the version used, the covariance may have different normalization terms τ; the set of indices m covers the range applicable in a particular case. The value s is called the time lag. Covariance applied to two different signals is known as cross-covariance (cross-correlation), and when X_i and X_j are the same signal the name auto-covariance (auto-correlation) is used.

In the frequency domain a function analogous to correlation is (ordinary) coherence. It compares common behavior of components of the signals at different frequencies. It is defined by means of elements of the spectral matrix $\mathbf{S}$

and depends on the signals X_i and X_j and on the frequency f. For simplicity a notation will be introduced: subscript indices of the function variable (ordinary coherence K) will correspond to the indices of the signals:

$$K(X_i, X_j, f) = K_{ij}(f) = \frac{S_{ij}(f)}{\sqrt{S_{ii}(f)S_{jj}(f)}}. \tag{4}$$

The modulus of ordinary coherence takes values in the [0,1] range. It describes the amount of in-phase components in both signals at the given frequency f(0 indicates no relation).

When a multivariate dataset consists of more than two signals, relations between them can be of a more complicated structure. Let us consider three signals X_1, X_2 and X_3 constituting a multivariate set. If X_1 influences X_2 and X_2 influences X_3 then signals X_1 and X_3 will be related with each other as well. However, contrary to relations of X_1 with X_2 and X_2 with X_3, the relation of channels X_1 and X_3 will not be direct because of the presence of intermediate signal X_2. Distinguishing between direct and indirect relations may play the crucial role in understanding the investigated system. To study relations within multi-($k > 2$)-variate datasets another functions were introduced.

Functions which help to decompose complex relations between signals and describe only direct ones are called partial functions. In multivariate systems partial functions identify only direct relations, with the influence of the rest of signals on that relations statistically removed. The partial coherence function $C_{ij}(f)$ describes the amount of in-phase components in signals i and j at frequency f while the part of the signals which can be explained by influence of a linear combination of the other signals is subtracted.

$$C_{ij}(f) = \frac{M_{ij}(f)}{\sqrt{M_{ii}(f)M_{jj}(f)}}. \tag{5}$$

Its modulus takes values within the [0, 1] range similar to ordinary coherence, but it is nonzero only when the relation between channel i and j is direct. M_{ij} is a minor (determinant) of $\mathbf{S}$ with i-th row and j-th column removed. After some algebraic manipulations C_{ij} can be expressed by elements of the inverse of $\mathbf{S}$: $d_{ij} = [S^{-1}]_{ij}$

$$C_{ij}(f) = (-1)^{i+j}\frac{d_{ji}(f)}{\sqrt{d_{jj}(f)d_{ii}(f)}}. \tag{6}$$

We may notice that in multichannel sets a signal can be simultaneously related with more than one signal. Multiple coherence $G_i(f)$ describes the amount of in-phase components in channel i common with any other channel of the set. It is given by the formula:

$$G_i(f) = \sqrt{1 - \frac{\det(\mathbf{S}(f))}{S_{ii}(f)M_{ii}(f)}}. \tag{7}$$

As in the case of other coherence functions, its modulus takes values within the [0, 1] range; its high value indicate the presence of a relation between the channel i and the rest of the set.

Partial coherences are especially useful to find a pattern of connections within sets of highly correlated signals. Example presented in Fig. 1 shows a result of coherence analysis of a 21-channels dataset of human scalp sleep EEG (see also Sect. 8). We see that each type of coherence forms a different pattern of connections. Multiple coherences (on the diagonal) are all high

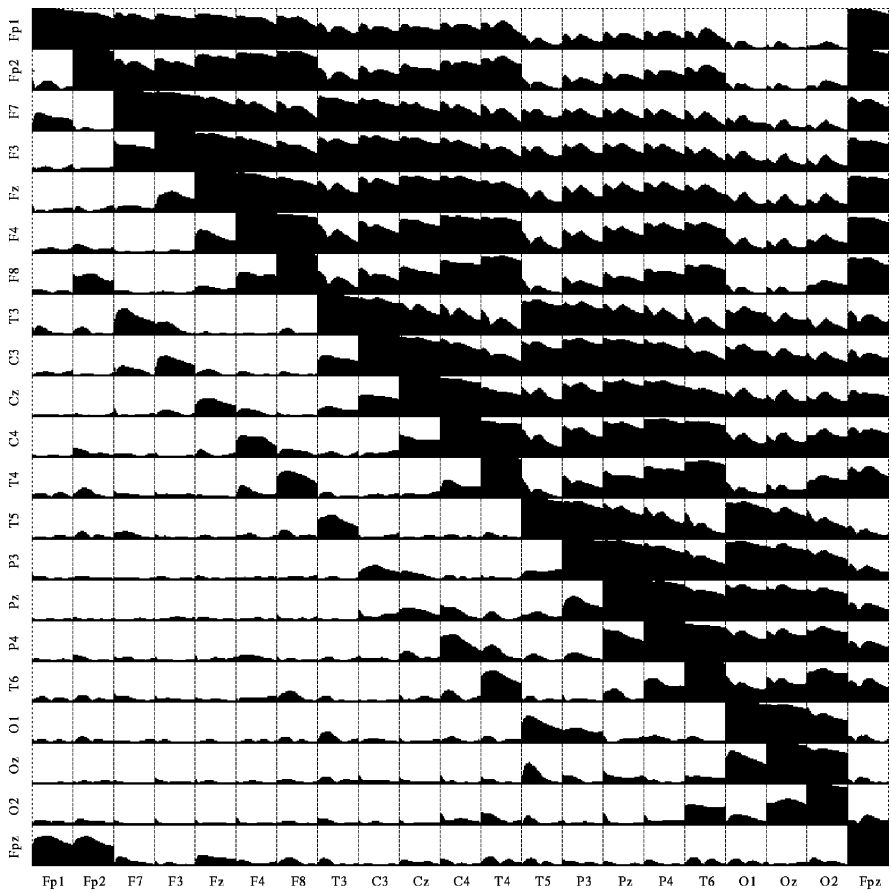


Fig. 1. Coherence analysis for a healthy human subject in sleep stage 2. Each box in the matrix is a graph of a coherence function connecting the channel marked below the relevant column and channel marked left of the relevant row. Frequency runs along the horizontal axes (0–30 Hz), the function value on the vertical axes (0–1). Ordinary coherences are plotted above the diagonal, multiple coherences on the diagonal and partial coherences below the diagonal of the matrix of graphs. Reprinted from (Kamiński et al. 1997) © 1997 with permission from International Federation of Clinical Neurophysiology

indicating strong relations of every channel with the rest of the set. While ordinary coherences (above the diagonal) are quite big and appear in certain groups, partial coherences have significant values only for few specific connections. Closer topographical analysis of those results revealed that the value of ordinary coherence depends mostly on the distance between electrodes. On the other hand, partial coherences mainly connect neighboring sites; for more distant locations they usually decrease quickly.

Coherence analysis is a popular and valuable tool in multichannel data analysis. When properly applied, it can quickly give insight into the connections pattern. Coherence results can be combined with other methods to get precise information about the network properties of the investigated system.

4 Parametric Modeling

In order to analyze data in the frequency domain, spectral quantities (e.g. $\mathbf{X}(f)$, $\mathbf{S}(f)$, etc.) have to be estimated. As said before, one very popular method is the Fourier transform, which gained popularity due to its ease of use. The Fast Fourier Transform (FFT) algorithm evaluates the spectral power of a signal in a fast and effective way. However, there are concurrent methods of spectral estimation, based on a parametric description of time series. In this approach a stochastic model of data generation is assumed. The model is fitted to the data resulting in a set of model parameters. The whole analysis is then conducted on the model parameters, not on the data samples.

The parametric approach has certain advantages over Fourier analysis (which belongs to the class of nonparametric methods, applied directly to the data). Parametric spectral estimates perform much better than nonparametric ones when applied to short data segments. In the Fourier approach the assumption is made that the time series are infinite or periodic. In practice, a finite and stochastic data epoch has to be analyzed. Finite stochastic datasets are then expressed as a multiplication of the signal by a window function, zeroing signal values outside the window. The window function induces distortions in the estimated spectra known as sidelobes. In the parametric approach, the validity of the model over the whole time scale is assumed, there is no need to introduce a window function and parametric spectra are smooth and free of sidelobe effects. More detailed discussions about these problems can be found in theoretical signal analysis textbooks (Kay 1988; Marple 1987) comparisons between Fourier methods and linear models can be found in (Isaksson et al. 1981; Blinowska 1994; Spyers-Ashby et al. 1998).

In biomedical data analysis practice two models are of primary importance: autoregressive (AR) and autoregressive-moving average (ARMA) models.

The multivariate autoregressive model (MVAR, MAR, VAR) is constructed by expressing $\mathbf{X}(t)$—a value of a (multivariate) process $\mathbf{X}$ at a time

t—by its past values taken with certain coefficients $\tilde{\mathbf{A}}$ and a (multivariate) white noise value $\mathbf{E}(t)$.

$$\mathbf{X}(t) = (X_1(t), X_2(t), \ldots, X_k(t))^{\mathrm{T}}, \mathbf{E}(t) = (E_1(t), E_2(t), \ldots, E_k(t))^{\mathrm{T}}$$
$$\mathbf{X}(t) = \sum_{j=1}^{p} \tilde{\mathbf{A}}(j)\mathbf{X}(t-j) + \mathbf{E}(t) \tag{8}$$

The $\tilde{\mathbf{A}}$ coefficients are the model parameters. The number p (of past samples taken into account) is called the model order.

Note that for given N time points of a k-variate process $\mathbf{X}$, we must estimate pk^2 parameters (p matrices $\tilde{\mathbf{A}}$ of size $k \times k$) from Nk data points.

Assuming $\tilde{\mathbf{A}}(0) = \mathbf{I}$ (the identity matrix) and $\mathbf{A}(j) = -\tilde{\mathbf{A}}(j)$, (8) can be rewritten in the form:

$$\mathbf{E}(t) = \sum_{j=0}^{p} \mathbf{A}(j)\mathbf{X}(t-j) \tag{9}$$

After transforming (9) into the frequency domain we obtain (Marple 1987):

$$\begin{aligned} \mathbf{E}(f) &= \mathbf{A}(f)\mathbf{X}(f) \\ \mathbf{X}(f) &= \mathbf{A}^{-1}(f)\mathbf{E}(f) = \mathbf{H}(f)\mathbf{E}(f) \\ \mathbf{H}(f) &= \left(\sum_{m=0}^{p} \mathbf{A}(m) \exp(-2\pi i m f \Delta t) \right)^{-1} \end{aligned} \tag{10}$$

(Δt is the data sampling interval). This equation leads to the observation that the signal in the frequency domain $\mathbf{X}(f)$ can be expressed as a product of $\mathbf{H}(f)$ and the white noise transform $\mathbf{E}(f)$. Because the spectral power of white noise is flat over frequency, the information about spectral properties of the process is contained in the matrix $\mathbf{H}$. This matrix is called the transfer matrix of the system.

The power spectrum of the signal is then given by

$$\mathbf{S}(f) = \mathbf{X}(f)\mathbf{X}^*(f) = \mathbf{H}(f)\mathbf{E}(f)\mathbf{E}^*(f)\mathbf{H}^*(f) = \mathbf{H}(f)\mathbf{V}\mathbf{H}^*(f) \tag{11}$$

where $\mathbf{V}$ denotes the input noise variance matrix (not dependent on frequency). The matrix $\mathbf{V}$ is evaluated from the data during the model fitting.

The autoregressive-moving average (ARMA) model of time series is described by:

$$\sum_{i=0}^{q} \mathbf{B}(i)\mathbf{E}(t-i) = \sum_{j=0}^{p} \mathbf{A}(j)\mathbf{X}(t-j) \tag{12}$$

where $\mathbf{B}(i)$ are parameters in addition to AR models; they are called a moving-average part. ARMA model can be viewed as an extension of AR model. Although the ARMA model is more universal than the AR model, it is rarely used in biomedical signal analysis. One reason is that the ARMA

model parameters estimation procedure is more complicated than the algorithms for AR model fitting. It often starts from an AR part estimation and then the MA parameters **B** are estimated separately. Second, it can be shown that the spectrum of the AR model can be fitted especially well to signals of a form of periodic components embedded in the noise (Franaszczuk and Blinowska 1985, Marple 1987). The rhythmic components are represented by peaks in the spectrum. Biomedical signals in general are of such type. A model with the **B** parameters can, in addition to modeling frequency peaks, describe dips in the spectrum. However, this signal feature is not typical for biomedical signals, so the ARMA model is seldom used.

Attempts to utilize AR models in biomedical signal processing date back to the 1960's (Achermann et al. 1994; Fenwick et al. 1969; Zetterberg 1969; Zetterberg 1973; Gersch 1970; Fenwick et al. 1971). AR modeling became popular with the wider accessibility to computers. Autoregressive models, especially in the multivariate version, are now quite often used, in particular in EEG and MEG analysis. Overviews of the linear modeling in applications to biomedical signals can be found in the literature (Jansen 1985; Kemp and Lopes da Silva 1991; Kelly et al. 1997; Kamiński and Liang 2005).

5 Causal Analysis

5.1 Defining Causal Estimators

Proper analysis of cross-relations in a multivariate dataset can provide information about causal relations between time series, for instance, sources of a signal can be identified. Before analyzing causal influences, causality for time series must be defined. The definition given by (Granger 1969), formulated originally for economic time series, recently became popular in biomedical data analysis. Its definition is expressed in terms of linear models of time series and can be easily applied to a parametric description of data.

Granger's original definition is based on predictability of time series. Let us assume that we try to predict the value of a process X_1 at a time t using p (an arbitrary number) past values of that process:

$$X_1(t) = \sum_{j=1}^{p} A_{11}(j) X_1(t-j) + \varepsilon(t) \tag{13}$$

We get a prediction error ε. If the prediction can be improved by adding to it some (q) values of another time series X_2 then we call X_2 causal for the X_1 series.

$$X_1(t) = \sum_{j=1}^{p} A'_{11}(j) X_1(t-j) + \sum_{j=0}^{q} A_{12}(j) X_2(t-j) + \varepsilon'(t) \tag{14}$$

The improvement of the prediction should be understood in a statistical sense, measured for instance by comparing the variances of the errors ε and ε'.

This definition can be extended to an arbitrary number (k) of signals. In that case we predict the signal $X_1(t)$ using all other available signals. That is to say if a signal X_m is causal for the X_1 the prediction error variance should be compared in two situations: when the signal X_m is either included or not included in the prediction

$$X_1(t) = \sum_{\substack{i=1 \\ (i \neq m)}}^{k} \sum_{j=1}^{p_i} A_{1i}(j) X_i(t-j) + \varepsilon''(t). \tag{15}$$

Historically, there were several attempts of defining various causal measures. Although the phase of coherence seems to be a good proposition for such a measure, in practice the ambiguity of phase values (which are defined modulo 2π) makes it difficult to utilize. Among proposed functions there were: various modifications and decompositions of coherences (like directed coherence (Baccalá and Sameshima 1998; Baccalá et al. 1998; Saito and Harashima 1981), analysis of feedback loops approach (Caines and Chan 1975; Gevers and Anderson 1981; Schnider et al. 1989), information theory measures (Kamitake et al. 1984; Saito and Harashima 1981; Gersch and Tharp 1976; Liang et al. 2001) and linear and nonlinear extensions to various versions of the correlation function (Chen et al. 2004; Freiwald et al. 1999; Chavez et al. 2003). In this chapter we will focus on methods based on and taking advantage of a parametric description of time series. Although applications of parametric (AR) modeling in causal relations analysis of biomedical data appear as early as the 1960's and 1970's (Whittle 1963; Gersch and Yonemoto 1977; Gersch 1972; Blinowska et al. 1981), it was often considered for bivariate systems rather than for an arbitrary number of signals. A truly multichannel measure, the Directed Transfer Function (DTF) was proposed in 1991 (Kamiński and Blinowska 1991). The DTF function operates in the frequency domain. Its construction is based on the elements of the transfer matrix $\mathbf{H}(f)$ of an AR model fitted to the whole multivariate system. The element $H_{ij}(f)$ can be related to the "amplitude" of the connection between input j and output i at frequency f. In the simplest (non-normalized) form the DTF is defined as

$$\theta_{ij}^2(f) = \left|H_{ij}(f)\right|^2 \tag{16}$$

Alternatively it can be calculated in a normalized form (Kamiński and Blinowska 1991):

$$\gamma_{ij}^2(f) = \frac{\left|H_{ij}(f)\right|^2}{\sum_{m=1}^{k} \left|H_{im}(f)\right|^2} \tag{17}$$

representing a ratio between the inflow to channel i from channel j to all inflows to channel i. The choice between the normalized and non-normalized version of DTF should be made according to a particular application.

The DTF, a measure constructed from elements of the transfer matrix $\mathbf{H}$, shows the total transmission between channels j and i, summed over all paths of the transmission. To indicate direct causal relations between channels a partial causal measure is needed. Partial Directed Coherence (PDC) was proposed by Baccalá and Sameshima (Baccalá and Sameshima 2001; Sameshima and Baccalá 1999). The PDC is constructed from $\mathbf{A}(f)$, elements of the Fourier transformed matrices of model coefficients $\mathbf{A}(t)$:

$$P_{ij}(f) = \frac{A_{ij}(f)}{\sqrt{\mathbf{a}_j^*(f)\mathbf{a}_j(f)}} \tag{18}$$

where $\mathbf{a}_j(f)$ denotes j-th column of the matrix $\mathbf{A}(f)$, the asterisk represents transposition and complex conjugate operation.

The normalization in PDC is different than in DTF; PDC shows a ratio between the outflow from channel j to channel i to all outflows from the source channel j. For comparison, we may rewrite the DTF definition (17) using a notation similar to (18):

$$\gamma_{ij}(f) = \frac{H_{ij}(f)}{\sqrt{\mathbf{h}_i^*(f)\mathbf{h}_i(f)}} \tag{19}$$

Here $\mathbf{h}_i(f)$ denotes i-th column of the matrix $\mathbf{H}(f)$.

Another measure of (partial) direct causal relations is proposed in (Korzeniewska et al. 2003). The direct DTF (dDTF) function is defined as a product of a modification of the DTF (H_{ij} in (20), with the denominator independent of frequency) with partial coherence.

$$H_{ij}^2(f) = \frac{\left|H_{ij}(f)\right|^2}{\sum\limits_f \sum\limits_{m=1}^{k} \left|H_{im}(f)\right|^2} \tag{20}$$

$$\chi_{ij}^2(f) = F_{ij}^2(f) C_{ij}^2(f)$$

The result has a nonzero value when both functions are nonzero which occurs when the given causal relation exists and is direct.

A comparison of DTF, PDC and dDTF applied to 21-channel human sleep EEG data is presented in Fig. 2 (see also Sect. 8 below). The data epoch was 20 s long; a MVAR model of order 4 was fitted and the respective functions were calculated. The values of the functions were integrated in 7–15 Hz range. The integrated functions are shown as topographic plots in form of arrows pointing toward the destination electrodes, the shade of gray represents the function value (black = the strongest). For clarity, each time only the 40 strongest flows are shown. In this case the alpha rhythm generators, known to be located in the back of the head, are active. The DTF pattern shows two main sources of activity around the electrodes P3 and P4, transmitting in various directions and especially to the front of the head. For dDTF the picture

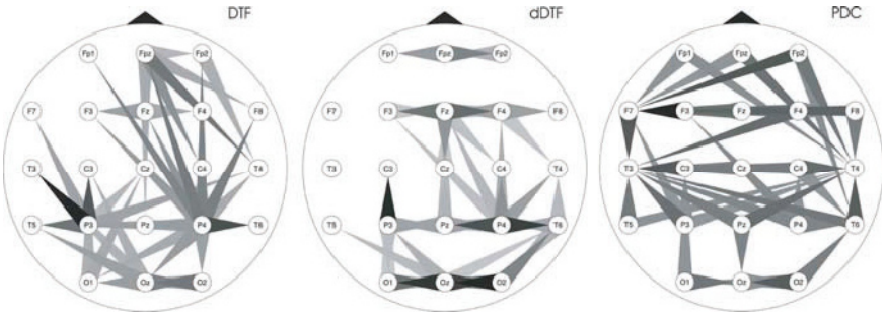


Fig. 2. Comparison of the DTF, dDTF and PDC functions applied to 21-channel human EEG sleep data. Flows are represented by arrows pointing from the source to the destination electrode. Intensity of the transmission is coded in shades of gray (black = the strongest). Only the 40 strongest flows are shown for each function. Reprinted from (Kuś et al. 2004) with permission (© 2004 IEEE)

is simpler. We see practically a subset of previously shown arrows where most of the longer distance connections do not exist anymore. On the PDC picture we find the main transmissions as indicated by dDTF and additional arrows indicating sinks of activity, as a result of the different normalization.

A comparison of DTF and PDC is also found in (Baccalá and Sameshima 2001). The relation between DTF, PDC and Granger causality is discussed in (Eichler 2006).

It should be noted that DTF does not give a value for the time delay between two signals. When the time relations are important other methods may perform better, e.g. the Hilbert transform in the time domain gives information about the instantaneous relative phase of a pair of signals which may be used for estimating the time delay.

5.2 Dynamic Phenomena

When estimating spectral properties of signals in the frequency domain, the data record length determines the statistical quality of the estimate. The model should be fitted to stationary records of data where the statistical properties of the signals do not change in time. Typically, the assumption of stationarity it is not valid for longer data records. When the record length is too short, the estimates become unreliable. Although linear modeling allows to perform spectral estimations even for very short data segments, nevertheless the limitation for a minimal record length is a serious problem in the analysis of dynamical processes. Data recordings of rapidly changing phenomena, like evoked potentials, cannot be split into long and stationary epochs, and other methods are required to treat them.

One solution to the above problem is to allow the model parameters to change in time. The big class of adaptive methods is based on this idea. Traditional adaptive algorithms include a progressive update of the model

parameters by extending the data epoch under consideration. Another technique is the recursive Kalman filter method (Arnold et al. 1998). Descriptions of those methods are given in the literature (Benveniste et al. 1990; Gath et al. 1992; Gersch 1987; Hesse et al. 2003; Möller et al. 2001; Schack et al. 1995; Schack and Krause 1995). In this chapter we will present yet another approach—an application of the short sliding window method to parametric spectral estimation. The procedure was proposed in (Ding et al. 2000) and applied in e.g. (Kamiński et al. 2001). It can be used when multiple repetitions of a process are available from recordings. In that case, the repetitions will be treated as different realizations of the same stochastic process. This assumption allows for utilization of all realizations in the estimation of spectral quantities. This way even data records of drastically reduced length (depending on the number of repetitions available) can still produce reliable spectral estimates. The whole nonstationary data record can be divided into short time windows, often short enough to treat the data inside them as stationary. If the window size is n, we must estimate pk^2 parameters from nkN_R data points (N_R is the number of repetitions). The bigger N_R, the shorter the window length n can be chosen. The actual decision about the window size must consider a balance between the time resolution of the results and the quality of the fit.

During model fitting the data covariance matrix is calculated for each repetition of the experiment. Then the averaged covariance matrix is calculated over all repetitions and this average is used to determine the parameters of a model within a given window. Note that the procedure does not involve data averaging. On the contrary, as a preprocessing step it is recommended to subtract the ensemble mean from all the repetitions; all future work is performed on the residual signals. This is done to fulfill the assumption of zero mean for the input signals. Omitting that step may result in unstable models for certain types of data. The details of data preprocessing and the procedure are described in (Ding et al. 2000; Kamiński et al. 2001). In Sect. 8 an example of the analysis will be presented.

The short-time DTF function (SDTF) (Kamiński et al. 2001) is an extension of the DTF allowing to trace the dynamics of a process. For each window a set of coherences and DTF functions can be calculated. By combining results for all the windows we obtain a picture of the dynamical changes of transmissions pattern between signals during the process. This approach assumes that all windows are processed with the same model order and the non-normalized DTF version is used in order to maximize comparability between the results from different windows.

6 Model Estimation

The first step in parametric analysis is fitting a model to the data. This is an essential step: the quality of the fit will determine our confidence in the results of the analysis. There are several issues which must be considered before starting the procedure.

The choice of the model is a fundamental decision. The AR model fits very well to data containing specific frequency rhythms with a noisy background. This description matches well the structure of EEG or MEG data, and in fact a big class of biomedical signals can be investigated using this formalism, including spike trains (as will be discussed later).

Another important step is to choose a proper data epoch: stationary and sufficiently long to produce a reliable fit (for the AR model we must estimate pk^2 parameters having Nk data points). It is difficult to give any precise limits, but it is safe to assume that we need several times more data points than parameters to be fit. If the data are nonstationary, a special approach for dynamic phenomena should be considered.

There is a wide range of publications concerning linear model fitting (Anderson 1984; Box et al. 1994; Hamilton 1994; Hayes 1996; Jenkins and Watts 1968; Kay 1988; Lutkepohl 1993; Marple 1987; Oppenheim and Schafer 1989; Priestley 1981; Proakis and Manolakis 1996) so there is no need to copy other handbooks here. Today a choice of a particular algorithm does not depend on calculation time anymore. However, certain algorithms may perform better when applied to certain types of data. Among different algorithms of parameters fitting the Yule-Walker (Marple 1987), Burg (Morf et al. 1978) and covariance (Schneider and Neumaier 2001) algorithms are often encountered. The Yule-Walker method is probably historically the first approach. The covariance matrix of the input signals is calculated and this matrix is the basis for solving a set of linear equations (called Yule-Walker equations) for the model parameters. The Burg algorithm is recursive, the data correlation matrix is not calculated. It provides high resolution for spectra. The covariance algorithm is useful for data with purely sinusoidal components over a stochastic background. Some comparisons of algorithms can be found in e.g. (Marple 1987) and other signal processing textbooks.

Typically, the algorithms assume data with zero mean which can be obtained by subtracting the temporal mean from every data channel. Sometimes it may be recommended to divide each channel by its temporal variance. Such a normalization can be useful if the data channels differ significantly in amplitude.

Next, the model order p must be selected. In practice, for a multivariate data the only possibility is to use one of the statistical criteria developed to find an optimal order. The optimal model order is a balance between obtaining a better fit by increasing the order (extending the number of past samples included) and keeping the order low to avoid artifacts which may appear in spectral estimates when the order is too big. Two criteria are particularly popular: AIC and FPE (Akaike 1974; Marple 1987). For instance, the AIC criterion is defined as:

$$\mathrm{AIC}(p) = n \log(\det(\mathbf{V})) + 2pk^2 \tag{21}$$

where n is the number of the data points, k is the number of channels and $\mathbf{V}$ is the noise variance matrix of the model. The criterion gives a numerical

value for a range of model orders ranging from 1 to a predefined maximum. Typically, the order for which a criterion reaches its first minimum is the optimal one.

Recently, Bayesian algorithms for optimal order selection and parameters fitting have been proposed (Box and Tiao 1992; Kitagawa and Gersch 1996; Penny and Roberts 2002).

7 Related Issues

7.1 Multivariate Versus Bivariate Time Series

In the previous sections some methods were referred to as "truly multichannel". This property cannot be attributed to methods applicable to a pair of channels only (pairwise or bivariate analysis). When a multivariate set of signals is analyzed in a bivariate manner, pair by pair, the whole structure of the set may not be revealed. Let us consider a simulation presenting the common source problem (Fig. 3).

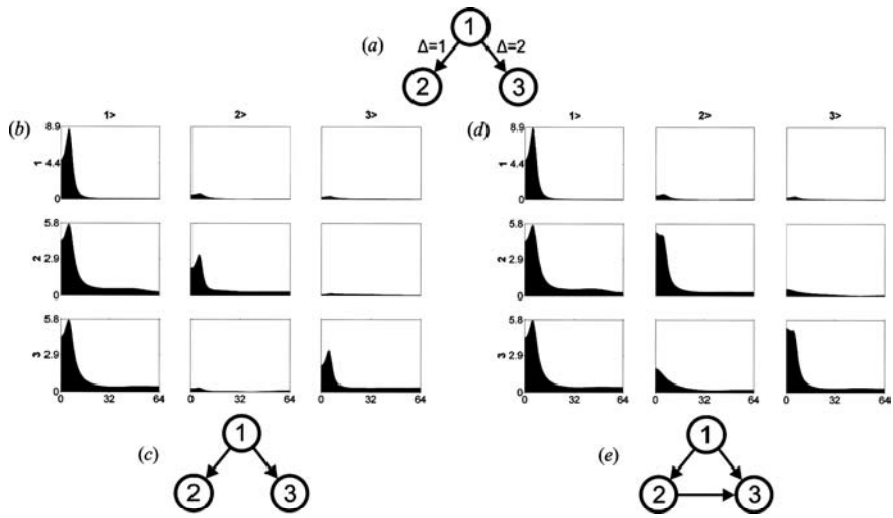


Fig. 3. DTF results for Simulation 1. (**a**) Scheme of simulated flows. Δ represents time delay in samples. (**b**) Multichannel DTF results. Each cell represents a (solid filled) graph of the function. Frequency is on the horizontal axis (0–64 Hz), the function value on the vertical axis. Each graph represents the function describing transmission from the channel marked above the relevant column to the channel marked left of the relevant row. On the diagonal of the matrix of graphs power spectra are shown. (**c**) Scheme of flows deducted from multichannel DTF (**d**) Pair-wise DTF results (organization like in *b*) (**e**) Scheme of flows deducted from pair-wise DTFs. Reprinted from (Kamiński 2005)

In this system channel 1 is a common source of signal for channels 2 and 3. Delays of the transmitted signals are different: one sample for $1 \rightarrow 2$ and two samples for $1 \rightarrow 3$ connections. This set of signals was analyzed twice using the same tool: the DTF function. In the first case DTF was applied to the whole 3-variate system. The second result was obtained applying DTF to every pair of channels of the system. In both cases we detect $1 \rightarrow 2$ and $1 \rightarrow 3$ transmissions. Additionally, in the bivariate case we detect a significant transmission $2 \rightarrow 3$ which originally did not exist in the system. This property of pairwise analysis may lead to improper conclusions, especially for more complicated situations. Such a more complex situation is presented in Fig. 4.

A record of human sleep EEG data (20 s, 21 channels) was analyzed twice using DTF calculated from one 21-variate model and 210 2-variate models (for every pair of channels). The patterns obtained for the connections are substantially different. In the case of bivariate analysis the alpha activity sources (located in the posterior region, visible in the first case around electrodes P3 and P4) are much less pronounced. Certain locations (F8) became sinks of activity. This effect is a direct consequence of a bivariate approach. Additional simulations showing possible pitfalls of bivariate analysis interpretation are given in (Kuś et al. 2004; Blinowska et al. 2004).

The fact that common sources of activity influence the interpretation of causality estimates is very important in experimental practice. Omitting a significant source of signal during a recording may lead to an improper

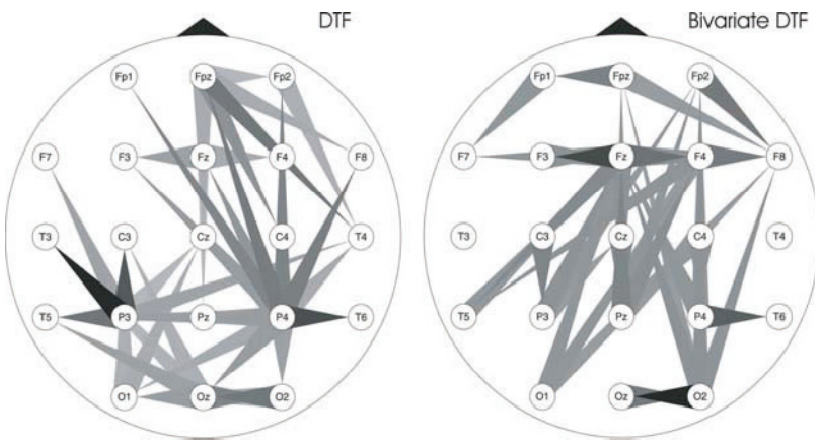


Fig. 4. Comparison of multichannel and pair-wise approach for estimating transmissions. DTF function was applied to a human EEG sleep data record twice–to the whole set of 21 channels simultaneously (left plot) and separately to every pair of channel (right plot). Flows are represented by arrows pointing from the source to the destination electrode. Intensity of transmission is coded in shade of gray (black = the strongest). Only the 40 strongest flows are shown for each case. Reprinted from (Kuś et al. 2004) with permission (© 2004 IEEE)

estimation of the connections pattern and false conclusions. If possible, all relevant sources of activity should be identified and included in the multivariate dataset.

7.2 Linear Versus Nonlinear Approaches

The formalism of linear parametric models describes the linear relations between signals. How can the linear approach be justified to process data generated by systems of inherently nonlinear nature (e.g. neurons)? The possibility of successful application of linear modeling depends on properties of the investigated system. Fortunately, there is evidence that many biomedical signals, like EEG, can be successfully treated with a linear approach and there is no need to apply a nonlinear procedure (Achermann et al. 1994, Veeramani et al. 2004, Blinowska and Malinowski 1991). Even during epilepsy, when nonlinear effects are most pronounced, we may still expect correct results (Stam et al. 1999, Pijn et al. 1991, Pijn et al. 1997). It has been demonstrated by several studies that linear methods can be used to solve some highly nonlinear problems such as the epileptogenic focus localization (Franaszczuk and Bergey 1998; Franaszczuk et al. 1994). Linear methods are quite simple to implement computationally, they have fewer restrictions about the input data characteristics than non-linear approach and can quickly give an insight into the nature of the problem. Nonlinear methods, which are typically more difficult to implement and more vulnerable to noisy signals, need to be used only when really necessary, possibly for a more detailed verification of linear analysis result.

It was shown in a separate study (David et al. 2004) that if there is a need for analyzing truly nonlinear relations, the risk of using a linear method is that we may not reveal all connections rather than obtaining completely false patterns (Franaszczuk et al. 1994; Zhu et al. 2003). However, a pure model situation is not common in practice. In reality, when recorded signals are intermixed from many different neuronal populations, the data effectively have a form of colored noise[1]. In such cases, there is a high risk that nonlinear methods may perform worse than linear analysis, producing inferior results.

7.3 Statistical Significance

In order to detect connections patterns correctly, statistical significance levels must be determined. Since quite often, especially for causal connection

[1] The pure white noise signal has its samples uncorrelated with each other for different time points. A spectrum of such signal is flat over the frequency. By a colored white noise we understand a signal which has this property disturbed – its power spectrum has a flat background but certain frequencies are "stronger". Of course it affects autocorrelation structure of such signal but it not necessarily affects relations of such signal with other signals.

estimators, it is difficult to give an analytical formulation for their statistical distributions, resampling statistics methods are applied.

Baseline levels are needed to reject insignificant connections. We may estimate them using a modification of the surrogate data algorithm proposed in (Theiler et al. 1992). The algorithm consists of the following steps: first, the data is Fourier transformed to the frequency domain. Then all relations between channels are removed by replacing all phases of the data by random numbers. The modified data is transformed back to time domain obtaining so-called surrogate data. The basic spectral properties of such signals are preserved (amplitudes of the transforms were unchanged), only the cross-channel relations have been removed. Analysis performed on the surrogate data should indicate zero cross-channel relations. By repeating the above procedure many times for (each time newly created) surrogate datasets we get a set of different estimates of coherences or DTF values for signals not related to each other. These estimates provide the statistical distribution of the baseline level.

Bootstrap technique (Efron 1979; Efron and Tibshirani 1993; Zoubir and Boashash 1998) can be used to evaluate the variance of SDTFs. In order to get the distribution of a function we simulate additional repetitions of the experiment: a randomly selected (out of all available) set of trials serves as a new input for the calculations. Again, we repeat the trial selection and functions calculation procedure many times. The set of estimates obtained for the repeated simulations provides the statistical distribution of the function value. Details about the procedures can be found in (Kamiński et al. 2001).

Recently, an analytical formula for significance levels for DTF has been proposed in (Eichler 2006).

7.4 Data Preparation

Multivariate analysis, and especially analysis of relations between channels, is vulnerable to proper data preprocessing. One must be very careful when preprocessing techniques are applied. Operations which introduce artificial correlations between channels, e.g. (Hjorth 1975) or laplacian, often performed for spatial "de-blurring", are unsuitable. During such the processing certain signals may become linear combinations of other signals. Similarly, the often used EEG practice of a "common average" reference electrode, where in fact a sum of all signals becomes the reference, should not be used here. The procedure mixes data from all the channels introducing strong correlations between all of them.

In experimental practice data are often filtered to enhance interesting frequency bands or remove artifacts. Any filters, analog or digital, may change phases of signals, possibly differently for different frequencies. Such filters, applied separately to different channels may distort estimates of relations between the channels in the frequency domain. Unfortunately, sometimes it is difficult to exactly know what operation is done on a measured signal within

a recording device by its firmware. When digital filtering is performed, all signals must be filtered identically and simultaneously or filters must be applied in both directions to cancel their phase shifts.

8 Examples of Applications

In this chapter a few typical examples of possible applications of the formalism debated above will be presented. Different practical situations will be considered allowing for a discussion of key issues related to techniques of multivariate data analysis using the linear modeling approach presented here.

8.1 Long and Stationary Data Epochs

In the previous sections various coherences and DTF results for human scalp sleep EEG data were already presented in Figs 1, 2 and 4. The data used for preparing those figures, were recorded as part of a bigger study of topographical patterns of connections during sleep (Kamiński et al. 1997). The data were collected from 21 electrodes placed on the human scalp according to 10–20 standard (Rechtschaffen and Kales 1968). The sampling frequency was 102.4 Hz. Whole night sleep data were recorded. Three independent experts scored a sleep stage for each 20 s epoch of data. To each continuous data segment of the same sleep stage a MVAR model was fitted. The optimal model order estimated by means of the AIC criterion was between 4 and 7. In this case (say, order 7) the number of fitted parameters is $21 \cdot 21 \cdot 7 = 3087$ and one second of input data contains $21 \cdot 102 = 2142$ numbers, so we need several seconds long epochs in order to get reliable estimates. In this study long and artifact free stationary data segments were available to the analysis; especially during the sleep stage 2 records were many dozens seconds long. From the model parameters power spectra, coherences and DTFs were calculated in the 0–30 Hz range. An example of the outcome for coherences is shown in Fig. 1. In order to better visualize the results the functions were integrated over interesting frequency ranges. The integrated values are shown as topographical plots of arrows (with grey shade indicating intensity). The typical DTF result (40 strongest flows) for sleep stage 2 is presented in Fig 2 (left).

An interesting comparison can be made between mapping and DTF. Although information from multiple channels is used to generate a map of spectral power, it cannot be considered a truly multichannel measure because no cross-channel information is utilized. In Fig. 5 mapping and DTF results for two EEG segments are shown (the segments are from recordings of EEG of human sleep stage 2, and wakefulness, respectively). For segment 1 we see a correspondence between the map and the DTF (a multichannel measure)—the maximum of spectral power falls off in the region identified as the source of the signal. For the second segment two sources were identified, transmitting signals diagonally to the front of the head. The maximal power on the map falls at the place where transmitted signals overlay each other.

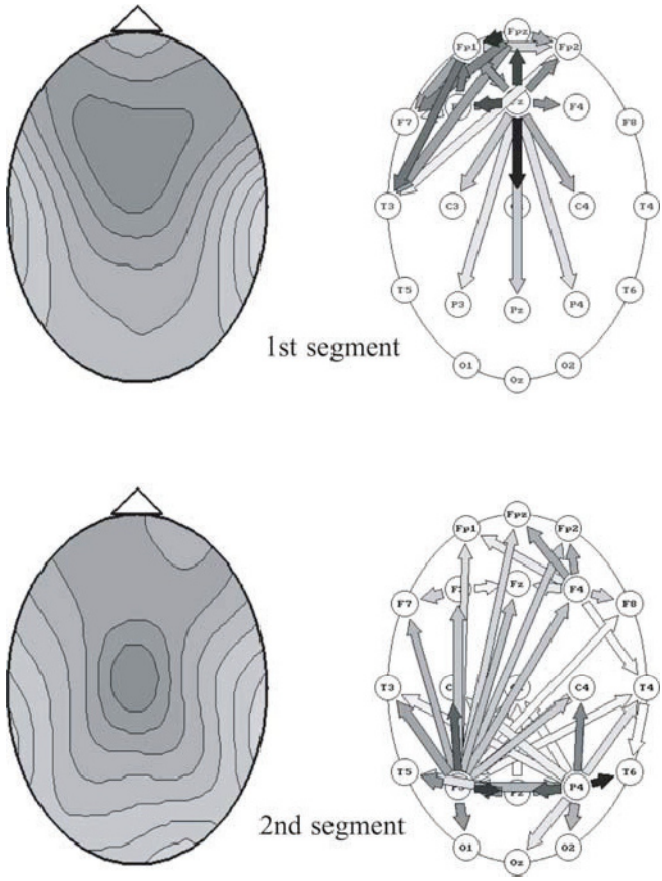


Fig. 5. Comparison of the DTF method with mapping. Two 20 s epochs of data were evaluated by both methods. Conventional maps represent total power in the 0–30 Hz frequency band, the DTF maps propagation in the same frequency band. The strength of flow is reflected by shades of gray of the arrows. Reprinted from (Kamiński et al. 1997) © 1997 with permission from International Federation of Clinical Neurophysiology

8.2 Dynamic Processes with Multiple Repetitions

Event related potentials are examples of dynamical phenomena where the character of the signals changes in time. The recorded data is not stationary and cannot be described by one linear model. However, when multiple repetitions of a brain response to the same stimulus are available, we may apply the short time window formalism. We treat the EEG signal from each repetition as a realization of the same stochastic process which has its origin in functional connections of brain structures. The signals are stochastic and different each time, but relations between them should be preserved as presumably generated by the same system every time. The following example

presents human ECoG recordings from Johns Hopkins University Hospital in Baltimore. The subjects were epilepsy patients with grids of electrodes implanted over various regions of the cortex. Some of the patients took part in a motor task study. They were required to clench their fist after a cue presentation on a screen (stimulus onset), sustain the contraction for three seconds, and release. Such tasks were repeated many times and 64-channel data were recorded and sampled at 1 kHz. For the further analysis seven seconds long epochs were selected from 3 s before to 4 s after the stimulus onset. For AR modeling, sets of 12 channels were selected according to their relevance to the task (a relative change of power during the task was measured and the placement over the hand motor cortex was considered).

To emphasize features in interesting frequency bands (beta: 20–30 Hz, gamma: 35–45 Hz) the data were lowpass frequency filtered and downsampled to 250 Hz. Later, bandpass filters were applied for the extraction of beta and gamma rhythms. The typical number of artifact-free trials was around 50. A time window of size 125 ms (31 samples) was used. For such a window the number of data points is $12 \cdot 31 \cdot 50 = 18600$; at the optimal (AIC criterion) model order 5 we must estimate only $12 \cdot 12 \cdot 5 = 720$ model parameters. Successive windows overlapped each other to get smoother results. Figure 6 presents SDTF results for the beta frequency range.

The figure is organized as a matrix of maps of the intensity of transmissions in time-frequency space. Placement of each map in the matrix determines which connection is described: from the electrode marked below the given column to the electrode marked left of the given row. Black color represents the strongest transmission, white color is indicating no transmission. Each map has a time scale -3–4 s on the horizontal axis (stimulus onset at time 0 marked by vertical lines) and frequency on the vertical scale from 20 to 50 Hz. We observe transmissions mostly in the 20–30 Hz range between specific electrodes. A characteristic feature is a rapid decrease of transmissions, occurring just after the stimulus onset, distinct especially for connections $35 \rightarrow 53$, $35 \rightarrow 51$, $53 \rightarrow 61$, $61 \rightarrow 53$, $43 \rightarrow 53$, $43 \rightarrow 45$ and others. This decrease of propagations can be connected with the effect of event-related desynchronization, known to occur for spectral power in the beta band. Additionally, we see that after the movement, certain transmissions reappear at slightly higher frequency and often at different locations (e.g. $28 \rightarrow 36$, $53 \rightarrow 35$, $61 \rightarrow 44$). This effect can be related to the beta rebound effect known from analysis of changes in power during movement tasks.

The matrix of SDTFs is a basis for many detailed analyses. Profiles of changes in specific frequencies at different time points can be investigated. As done previously, integrated values of SDTF can be used for topographical presentation of the results in form of arrows. In this situation, a movie can be created visualizing dynamic changes of transmission patterns. As an example, a frame from a movie showing transmissions in the beta range for 1.7 s before stimulus onset is given in Fig. 7. More about this experiment was published in (Kamiński et al. 2005), see also (Crone et al. 1998a, Crone et al. 1998b).

Fig. 6. SDTF results (transmissions) in the beta range (20–30 Hz) for a set of 12 electrodes for patient 1. In each small panel SDTF as a function of time (horizontal axis, from −3 to 4 s) and frequency (vertical axis, from 20 to 50 Hz) is shown. For the sake of clarity, a bigger legend is given only for the bottom left panel; settings are the same for all the panels. The flow of activity takes place from the electrode marked under the column to the electrode marked left of the relevant row. Black color corresponds to the highest intensity, white to the lowest. The value of the transmission grayscale is the same for all the panels. Time zero is marked by vertical lines. Reprinted from (Kamiński et al. 2005) with permission

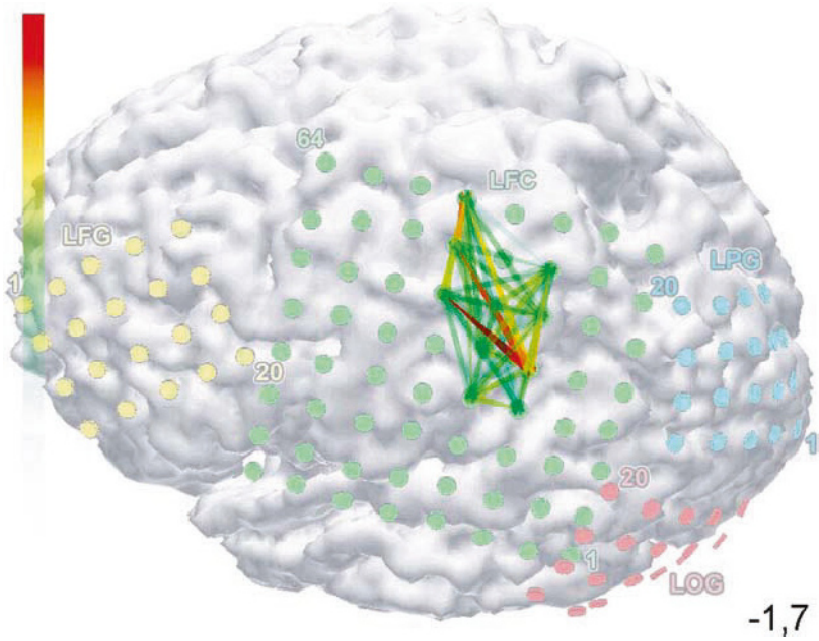


Fig. 7. Propagations in beta range (20–30 Hz) for a set of 10 electrodes for 1.7 s before stimulus onset. Directions are shown as triangle-shapeds arrow pointing from the source to the destination electrodes. Intensity of transmissions is coded in color and transparency of arrows (scale bar on the left of the picture). Reprinted from (Kamiński et al. 2005) with permission

8.3 Hybrid Spike Train-LFP Dataset

Spike train data have the form of a point process. Such type of data does not conform to the stochastic time series description. The problem of spectral analysis of spike trains was considered important as early as in 1971 (French and Holden 1971). However, after some preprocessing, even spike train signals can be analyzed using AR modeling (Brillinger 1978).

In the experiment signals were recorded in the brain structures of a rat (Kamiński and Kocsis 2003). Three sites related to generation of activity in the theta (3.7–5.6 Hz) range were chosen: septum (SEP), supramammilary nucleus (SUM) and hippocampus (HIP). The role of the SUM, a structure reciprocally connected with two other structures, was investigated during theta rhythm generation. In the hippocampus local field potentials (LFP) were recorded while in the SUM the specific activity of selected single neurons was measured. Sampling frequency of the LFP was 250 Hz. To prepare the data for AR modeling, the spike trains were lowpass filtered at a 12.5 Hz cutoff frequency. The stochastic component was added to the filtered signals in form of white noise of small amplitude (10% of the signal). The data for different stages of processing are shown in Fig. 8.

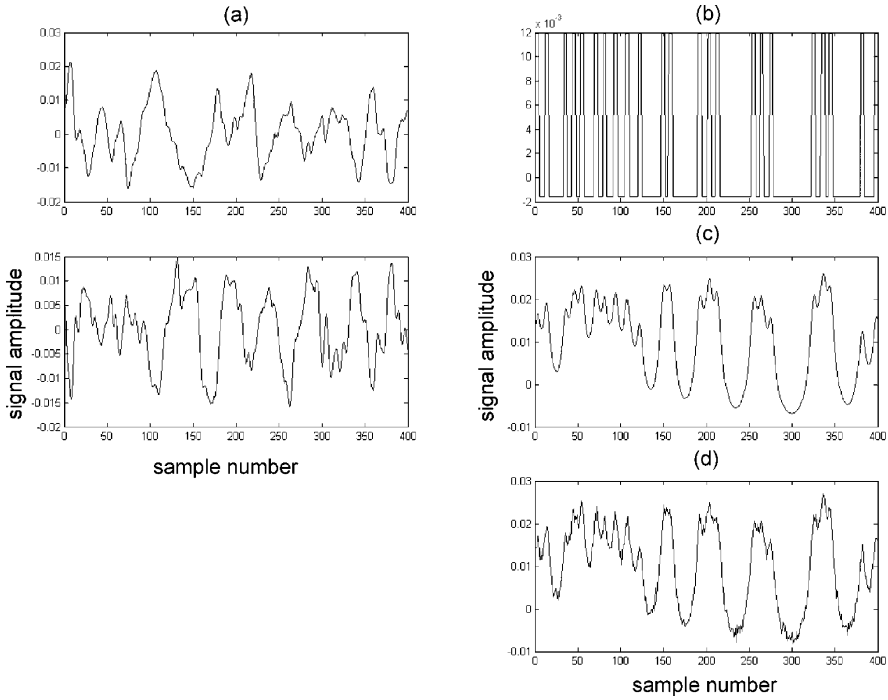


Fig. 8. Examples of signals investigated in the study. (**a**) two channels of LFP with theta activity recorded from the hippocampus; (**b**) spike train recorded from a neuron located in SUM; (**c**) lowpass filtered spike train from section c; (**d**) lowpass filtered spike train (from section c) with 10% of a noise added. Reprinted from (Kamiński and Liang 2005) with permission from Begell House, Inc., © 2005

Changes in theta rhythm generation occurred during pinching the tail of a rat. Interesting fragments of data including the tail pinch period were selected for causal relations analysis. Although there was only one experiment with no repetitions, the short time window methodology was used to visualize changes of transmissions in different situations. The sliding window has to be long enough to fulfill requirements about statistical properties of the estimates and, moreover, it must contain several spikes. On the other hand, the window should be short enough to capture the dynamics of the changes. Finally, a 500 samples (2 s) long window was chosen. Successive windows were shifted with 10 samples of overlap between each other. Example of SDTF results for SUM and HIP is shown in Fig. 9.

The SDTF calculated for several neurons revealed a similar trend: although both influences SUM → HIP and HIP → SUM are present, the HIP → SUM transmission is stronger and changes its character during the tail pinch. The description of the whole experiment can be found in (Kocsis and Kamiński 2006).

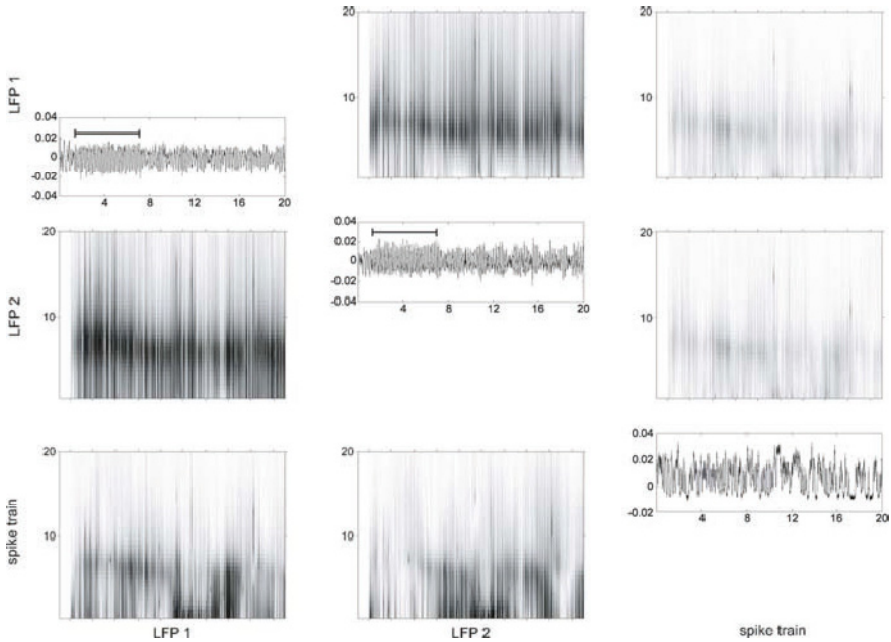


Fig. 9. Example of SDTF functions for set of: two LFP traces recorded in the hippocampus (LFP 1, LFP 2) and filtered spike trains recorded from a neuron located in SUM. On the diagonal the data time courses are shown. Time on horizontal axis. Tail pinch from 1 do 7 s (marked by bars above data time courses). In grey panels frequency in Hz on vertical axis. Convention similar to the scheme used in Fig. 6. Reprinted from (Kamiński and Liang 2005) with permission from Begell House, Inc., © 2005

The above methodology is suitable for spike trains containing a large number of spikes. When the spikes are sparsely distributed, better results could be obtained after converting the spike train into time series of the instantaneous firing rate (Zhu et al. 2003).

9 Summary and Discussion

In this chapter an introduction to basic problems concerning multivariate analysis was presented. To interpret the informational content of a multivariate dataset correctly we must take into account both auto-quantities, describing properties of every channel separately, and cross-quantities describing common network properties of the whole set (inter-relations between signals). Correlation and coherence functions are popular and useful examples of cross-functions. One must remember that with the number of channels increasing, more possibilities arise for relations between signals. Conducting the research in the frequency domain, partial coherences allow to distinguish

between simple (direct) and more complex connections. Although coherence functions can be estimated from data in a nonparametric way, the parametric approach is an interesting alternative providing several advantages. In parametric analysis a linear stochastic model is assumed as the generation scheme of the data. From the model parameters power spectra, coherences and other functions can be estimated. MVAR spectra are smooth and can be estimated reliably from short data epochs. Moreover, a multivariate model can be fitted simultaneously to all signals as a single system, therefore we get truly multivariate estimates, free of certain interpretation pitfalls. When applying a causal estimator it is important to include the whole set of relevant channels. If the signals are not treated as one system and measures are applied pair-wise the results may be misleading or at best difficult to interpret. Another advantage of parametric models is their natural ability of defining a causal multivariate estimator describing transmissions of signals between data channels.

The most popular linear models are the AR and ARMA models. The AR model is especially well suited for describing damped sinusoidal signals over a noisy background and is valid for many biomedical signals. AR model equations in the frequency domain have the form of a transfer matrix operating on white noise. The transfer matrix is the basis for constructing the Directed Transfer Function, a multivariate estimator of causal relations between channels. Partial causal estimators were also proposed in the literature, they include the PDC and dDTF functions presented above.

Linear modeling is not in contradiction to nonlinear methods. In doubtful cases, the linear approach can be applied to get a first insight into the connections of the system. The selected relations can be further investigated using nonlinear analysis tools. Similarly, if a bivariate method should be applied, it is recommended to process the whole dataset in the multivariate mode and then select interesting relations on the basis of multivariate results for further bivariate analysis.

As shown in the presented examples, linear modeling can be applied to various types of multivariate data differing in length, number of channels and overall character. The examples were selected in order to cover methodological issues discussed before. In the presented examples sleep EEG, task related ECoG, rat LFP and spike train data were analyzed. Of course the list can be extended to another types of signals (e.g. fMRI (Harrison et al. 2003) or cardiovascular signals (Korhonen et al. 1996)), making the parametric approach a very useful tool in various applications, not only in biomedical research.

Acknowledgments

The author thanks Prof. K. J. Blinowska from Warsaw University for helpful discussion.

References

Achermann P, Hartmann R, Gunzinger A, Guggenbühl W, Borbély AA. All night sleep and artificial stochastic control have similar correlation dimension. Electroenceph Clin Neurophys 1994; 90:384–387.

Akaike H. On the use of a linear model for the identification of feedback systems. Ann Inst Stat Math 1968; 20:425–40.

Akaike H. A new look at statistical model identification. IEEE Trans Autom Control 1974; 19:716–23.

Anderson TW. An introduction to multivariate statistical analysis. John Wiley & Sons, 1984.

Arnold M, Miltner WHR, Bauer R, Braun C. Multivariate autoregressive modeling of nonstationary time series by means of Kalman filtering. IEEE Trans Biomed Eng 1998; 45:553–62.

Baccalá LA, Sameshima K. Directed coherence: a tool for exploring functional interactions among brain structures. In: Methods for neural ensemble recordings. CRC Methods In Neuroscience series, CRC Press, Boca Raton, 1998.

Baccalá LA, Sameshima K. Overcoming the limitations of correlation analysis for many simultaneously processed neural structures. Progress Brain Res 2001; 130:33–47.

Baccalá LA, Sameshima K. Partial directed coherence: a new concept in neural structure determination. Biol Cybern 2001; 84:463–74.

Baccalá LA, Sameshima K, Ballester G, Do Valle AC, Timo-Iaria C. Studying the interaction between brain structures via directed coherence and Granger causality. App Sig Proc 1998; 5:40–48.

Benveniste A, Metivier M, Priouret P. Adaptive algorithms and stochastic approximations. Springer-Verlag, 1990.

Blinowska KJ. New trends in electrophysiological signals parameterization and feature extraction. Technol Health Care 1994; 2:93–110.

Blinowska KJ, Czerwosz LT, Drabik W, Franaszczuk PJ, Ekiert H. EEG data reduction by means of autoregressive representation and discriminant analysis procedures. Electroenceph Clin Neurophys 1981; 51:650–8.

Blinowska KJ, Kuś R, Kamiński M. Granger causality and information flow in multivariate processes. Phys Rev E 2004; 70:050902, Virt J Biol Phys Res 2004; 8(11).

Blinowska KJ, Malinowski M. Non-linear and linear forecasting of the EEG time series. Biol Cybern 1991; 66:159–165.

Box G, Jenkins GM, Reinsel G. Time series analysis: forecasting and control. Prentice Hall, 1994.

Box GEP, Tiao GC Bayesian inference in statistical analysis. John Wiley, NY, 1992.

Brillinger DR. Comparative aspects of the study of ordinary time series and point processes. Devel Stat 1978; 5:33–133.

Caines PE, Chan CW. Feedback between stationary stochastic processes. IEEE Trans Autom Control 1975; 20:498–508.

Chavez M, Martinerie J, Le Van Quyen M. Statistical assessment of nonlinear causality: application to epileptic EEG signals. J Neurosci Meth 2003; 124:113–28.

Chen Y, Rangarajan G, Feng J, Ding M. Analyzing multiple nonlinear time series with extended Granger causality. Phys Lett A 2004; 324:26–35.

Crone NE, Miglioretti DL, Gordon B, Sieracki JM, Wilson MT, Uematsu S, Lesser RP. Fuctional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. I. Alpha and beta event-related desynchronization. Brain 1998a; 121:2271–99.

Crone NE, Miglioretti DL, Gordon B, Lesser RP. Fuctional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. Brain 1998b; 121:2301–15.

David O, Cosmelli D, Friston KJ. Evaluation of different measures of functional connectivity using a neural mass model. NeuroImage 2004; 21:659–73.

Ding M, Bressler SL, Yang W, Liang H. Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. Biol Cybern 2000; 83:35–45.

Eichler M. On the evaluation of information flow in multivariate systems by the directed transfer function. Biol Cybern 2006; 94(6): 469–82.

Efron B. Bootstrap methods: another look at the jackknife. Ann Stat 1979; 7:1–26.

Efron B, Tibshirani RJ. An introduction to the bootstrap. Chapman & Hall, 1993.

Fenwick PBC, Mitchie P, Dollimore J, Fenton GW. Application of the autoregressive model to EEG analysis. Agressologie 1969; 10:553–64.

Fenwick PBC, Mitchie P, Dollimore J, Fenton GW. Mathematical simulation of the electroencephalogram using an autoregressive series. Int J Bio-Med Comput 1971; 2:281–307.

Franaszczuk PJ, Bergey GK. Application of the directed transfer function method to mesial and lateral onset temporal lobe seizures. Brain Topogr 1998; 11:13–21.

Franaszczuk PJ, Bergey GK, Kamiński M. Analysis of mesial temporal seizure onset and propagation using the directed transfer function method. Electroenceph Clin Neurophys 1994; 91:413–27.

Franaszczuk PJ, Blinowska KJ. Linear model of brain electrical activity – EEG as a superposition of damped oscillatory modes. Biol Cybern 1985; 53:19–25.

Freiwald WA, Valdes P, Bosch J, Biscay R, Jimenez JC, Rodriguez LM, Rodriguez V, Kreiter AK, Singer W. Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex. J Neurosci Meth 1999; 94:105–19.

French AS, Holden AV. Alias-free sampling of neuronal spike trains. Biol Cyb 1971; 8(5):165–171.

Gath I, Feuerstein C, Pham DT, Rondouin G. On the tracking of rapid dynamic changes in seizure EEG. IEEE Trans Biomed Eng 1992; 39:952–8.

Gersch W. Causality or driving in electrophysiological signal analysis. Math Biosci 1972; 14:177–96.

Gersch W. Spectral analysis of EEG's by autoregressive decomposition of time series. Math Biosci 1970; 7:205–22.

Gersch W. Non-stationary multichannel time series analysis. In: Handbook of electroencephalography and clinical neurophysiology; Vol. 1: Methods of analysis of brain electrical and magnetic signal. Amsterdam: Elsevier, 261–296, 1987.

Gersch W, Tharp BR. Spectral regression-amount of information analysis of seizures in humans. In: Kellaway P, Petersen I, editors. Quantitative Analytic Studies in Epilepsy. New York: Raven Press, 509–532, 1976.

Gersch W, Yonemoto J. Parametric time series models for multivariate EEG analysis. Comput Biomed Res 1977, 10:113–25.

Gevers MR, Anderson BDO. Representations of jointly stationary stochastic feedback processes. Int J Control 1981; 33:777–809.

Granger CWJ. Investigating causal relations in by econometric models and cross-spectral methods. Econometrica 1969; 37:424–38.

Hamilton JD. Time series analysis. Princeton, NJ: Princeton Univ. 1994.

Harrison L, Penny WD, Friston K. Multivariate autoregressive modeling of fMRI time series. NeuroImage 2003; 19:1477–91.

Hayes MH. Statistical digital signal processing and modeling. New York: Wiley & Sons, 1996.

Hesse W, Möller E, Arnold M, Schack B. The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies. J Neurosci Meth 2003; 124:27–44.

Hjorth B. An on-line transformation of EEG scalp potentials into orthogonal source EEG channels. Electroenceph Clin Neurophys 1975; 39:526–30.

Isaksson A, Wennberg A, Zetterberg LH. Computer analysis of EEG signals with parametric models. Proc IEEE 1981; 69:451–61.

Jansen BH. Analysis of biomedical signals by means of linear modeling. CRC Crit Rev Biomed Eng 1985; 12(4):343–92.

Jenkins GM, Watts DG. Spectral analysis and its applications. SF: Holden-Day, 1968.

Kamiński M. Determination of transmission patterns in multichannel data. Philosophical Transactions of The Royal Society B, 2005, 360: 947–952.

Kamiński M, Blinowska KJ. A new method of the description of the information flow in the brain structures. Biol Cybern 1991; 65:203–10.

Kamiński M, Blinowska KJ, Szelenberger W. Investigation of coherence structure and EEG activity propagation during sleep. Acta Neurobiol Exp 1995; 55:213–9.

Kamiński M, Blinowska KJ, Szelenberger W. Topographic analysis of coherence and propagation of EEG activity during sleep and wakefulness. Electroenceph Clin Neurophys 1997; 102:216–27.

Kamiński M, Ding M, Truccolo W, Bressler S. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. Biol Cybern 2001; 85:145–57.

Kamiński M, Kocsis B. Direction of the theta rhythmic drive between neurons in the supramammillary nucleus (SUM) and the septohippocampal system in urethane anesthetized rats. Neurosci Abs 2003; 29:938.13.

Kamiński M, Liang H. Causal influence: advances in neurosignal analysis. Critical Reviews in Biomedical Engineering, 2005, 33(4): 347–430.

Kamiński M., Zygierewicz J, Kus R, Crone N. Analysis of multichannel biomedical data. Acta Neurobiologiae Experimentalis, 2005, 65(4): 443–452.

Kamitake T, Harashima H, Miyakawa H. A time-series analysis method based on the directed transinformation. Electronics and Communications in Japan 1984; 67-A(6).

Kay SM. Modern spectral estimation. Englewood Cliffs, NJ: Prentice Hall, 1988.

Kelly EF, Lenz JE, Franaszczuk PJ, Truong YK. A general statistical framework for frequency-domain analysis of EEG topographic structure. Comput Biomed Res 1997; 30:129–64.

Kemp B, Lopes da Silva FH. Model-based analysis of neurophysiological signals. In: Weitkunat R, editor. Digital biosignal processing, Amsterdam: Elsevier, 129–155, 1991.

Kitagawa G, Gersch W. Smoothness priors analysis of time series. Lecture Notes in Statistics, NY: Springer, 116, 1996.

Kocsis B., Kamiński M. Dynamic changes in the direction of the theta rhythmic drive between supramammillary nucleus and the septohippocampal system. Hippocampus 2006; 16(6): 531–40.

Korhonen I, Mainardi L, Baselli G, Bianchi A, Loula P, Carrault G. Linear multivariate models for physiological signal analysis: applications. Comput Meth Prog Biomedic 1996; 51:121–30.

Korzeniewska A, Mańczak M, Kamiński M, Blinowska KJ, Kasicki S. Determination of information flow direction between brain structures by a modified Directed Transfer Function method (dDTF). J Neurosci Meth 2003; 125:195–207.

Kuś R, Kamiński M, Blinowska KJ. Determination of EEG activity propagation: pair-wise versus multichannel estimate. IEEE Trans Biomed Eng 2004; 51: 1501–10.

Liang H, Ding M, Bressler SL. Temporal Dynamics Of Information Flow In The Cerebral Cortex. Neurocomputing 2001; 38–40:1429–35.

Lutkepohl H. Introduction To Multiple Time Series Analysis. NY: Springer, 1993.

Marple SL. Digital Spectral Analysis With Applications Prentice-Hall Signal Processing Series. New Jersey: Simon & Schuster, 1987.

Möller E, Schack B, Arnold M, Witte H. Instantaneous multivariate EEG coherence analysis by means of adaptive high-dimensional autoregressive models. J Neurosci Meth 2001; 105:143–58.

Morf M, Vieira A, Lee D, Kailath T. Recursive multichannel maximum entropy spectral estimation. IEEE Trans Geosci Electronics 1978; 16:85–94.

Oppenheim AV, Schafer RW. Discrete-time signal processing. Prentice Hall. NJ: Englewood Cliffs, 1989.

Penny WD, Roberts SJ. Bayesian multivariate autoregressive models with structured priors. IEE Proc Vision Image Sig Proc 2002; 149:33–41.

Pijn JPM, Van Neerven L, Noest A, Lopes da Silva FH. Chaos or noise in EEG signals: dependence on state and brain site. Electroenceph Clin Neurophys 1991; 79:371–381.

Pijn JPM, Velis DN, van der Heyden MJ, DeGoede J, van Veelen CWM, Lopes da Silva FH. Nonlinear dynamics of epileptic seizures on basis of intracranial EEG recordings. Brain Topogr 1997; 9:249–270.

Priestley MB. Spectral analysis and time series. Academic Press, 1981.

Proakis JG, Manolakis DG. Digital signal processing: principles, algorithms and applications. Prentice Hall. NJ: Englewood Cliffs, 1996.

Rechtschaffen A, Kales A. A manual: standardized terminology, techniques and scoring system for sleep stages of human subjects. Brain Information Service. Brain Research Institute, UCLA, 1968.

Saito Y, Harashima H. Tracking of information within multichannel record: causal analysis in EEG. In Yamaguchi N, Fujisawa K, editors. Recent Advances in EEG and EMG Data Processing. Amsterdam: Elsevier, 1981, 133–146.

Sameshima K, Baccalá LA. Using partial directed coherence to describe neuronal ensemble interactions. J Neurosci Meth 1999; 94:93–103.

Schack B, Grieszbach G, Arnold M, Bolten J. Dynamic cross-spectral analysis of biological signals by means of bivariate ARMA process with time-dependent coefficients. Med & Biol Eng & Comput 1995; 33:605–10.

Schack B, Krause W. Dynamic power and coherence analysis of ultra short-term cognitive process, a cognitive study. Brain Topogr 1995; 9:127–36.

Schneider T, Neumaier A. Estimation of Parameters and Eigenmodes of Multivariate Autoregressive Models. ACM Trans Math Soft 2001; 27:27–57, 58–65.

Schnider SM, Kwong RH, Lenz FA, Kwan HC. Detection of feedback in the central nervous system using system identification techniques. Biol Cybern 1989; 60:203–12.

Spyers-Ashby JM, Bain PG, Roberts SJ. A comparison of fast Fourier transform (FFT) and autoregressive (AR) spectral estimation techniques for the analysis of tremor data. J Neurosci Meth 1998; 83:35–43.

Stam C, Pijn JPM, Suffczyski P, Lopes da Silva FH. Dynamics of the human alpha rhythm: evidence for non-linearity? Clin Neurophys 1999; 110:1801–1813.

Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer D. Testing for nonlinearity in time series: the method of surrogate data. Physica D 1992; 58:77–94.

Veeramani B, Narayanan K, Prasad A, Iasemidis LD, Spanias AS, Tsakalis K. Measuring the direction and the strength of coupling in nonlinear systems – a modeling approach in the state space. IEEE Sig Proc Lett 2004; 11:617–620.

Whittle P. On the fitting of multivariate autoregressions and the approximate factorization of a spectral matrix. Biometrika 1963; 50:129–34.

Zetterberg L. Estimation of parameters for a linear difference equation with application to EEG analysis. Math Biosci 1969; 5:227–75.

Zetterberg L. Experience with analysis and simulation of EEG signals with parametric description of spectra. In: Kellaway P, Petersen I, editors. Automation of Clinical Electroencephalography, NY: Raven Press, 161–201, 227–234, 1973.

Zhu L, Lai Y-C, Hoppensteadt FC, He J. Probing changes in neural interaction during adaptation. Neural Comput 2003; 15:2359–77.

Zoubir AM, Boashash B. The bootstrap and its application in signal processing. IEEE Sig Proc Mag 1998; 15:56–76.

Beamforming and Its Applications to Brain Connectivity

Armin Fuchs

Center for Complex Systems and Brain Sciences, Department of Physics,
Florida Atlantic University, Boca Raton, FL 33487
fuchs@ccs.fau.edu

In this chapter we show how beamforming, an analysis procedure for EEG and MEG data sets that is becoming increasingly popular, can be used to obtain insight about functional connectivity between brain regions. To this end we introduce a parameterization of cortical surfaces based on output from the software package *Freesurfer* and restrict the potential current sources to the cortical gray matter. We create a data set assuming two brain areas being active with different time dependencies of the activation. These activation patterns are then reconstructed using beamforming procedures with a variety of parameters demonstrating the dependencies of the reconstructed patters of activity on these parameters. Finally, we apply two statistical techniques, namely coherence and Granger causality to the data set, showing that the data can be understood by assuming a one-directional connection between the two active brain regions.

1 Introduction

In recent years an analysis procedure called beamforming has become increasingly popular for the analysis of non-invasively recorded electrophysiological data sets like EEG (electroencephalogram) and MEG (magnetoencephalogram). The original ideas can be found in the engineering literature of the 1970's (Frost III 1972, Borgiotti and Kaplan 1979) (see also Van Veen and Buckley 1988) before they were applied to brain imaging roughly 25 years later (Sekihara 1996, Van Veen et al. 1997, Robinson and Vrba 1999). The goal of beamforming is to use an array of sensors and combine the signals recorded at individual sites to increase the signal/noise ratio or resolution, and to focus the entire array on a certain region in space. The best known example for this technique is found in astronomy where many radio telescopes that can be distributed all over the earth behave as a virtual single antenna. In brain imaging, arrays of up to about 250 electrodes (EEG) or SQuID (superconducting quantum interference device) sensors (MEG) are used to measure the electric

potential on the scalp surface or the tiny magnetic fields produced by the electric currents flowing inside the brain, respectively. When beamforming is applied to electrophysiological recordings the intention is to find the magnitudes, locations and directions of these currents, in short, detect the neurally active regions of the brain under a certain task. Various flavors of beamforming algorithms can be found in the literature and we will discuss their similarities and differences later in this chapter. In most applications beamformers are used to scan regions of interest or the entire brain volume on a voxel by voxel basis (Sekihara 1996, Van Veen et al. 1997, Robinson and Vrba 1999). More recently a slightly different approach has been developed which applies anatomical constraints to the locations and directions of potential cortical sources (Fuchs 2002, Hillebrand and Barnes 2003). Specifically, the locations of neural activity are restricted to the cortical gray matter and the directions of the currents are assumed to be perpendicular to the cortical surface. The reasons for these constraints on location and direction are twofold: first, in the cortical gray matter pyramidal cells form columns which are orthogonal to the folded surface and the current flow is along these columns (Abeles 1991). Second, the columns are bundled together to form so-called macrocolumns of about 10^5 neurons acting coherently (Braitenberg and Schüz 1991). It has been estimated that several tens of thousands of simultaneously active cells are necessary to produce a signal that can be picked up by EEG or MEG (Nunez 1981) and the only part of the brain where this is the case is the cortical gray matter[1].

We will proceed through this chapter in four steps: first, we will show how MRI scans and the locations of sensors of a magnetometer used to record MEG can be coregistered to a common coordinate system, and how the cortical surfaces can be extracted and parameterized. Second, we will create a dataset by assuming neural activity in two distinct regions along opposite walls of the central sulcus and calculate the spatiotemporal patterns that would be measured by the MEG system. Third, we will use this dataset and apply beamformers to reconstruct the activity in order to show how certain parameters affect the quality of performance. Finally, fourth, we will demonstrate how the statistical measures known as coherence and Granger causality allow us to identify functional connectivity between brain regions including the direction of influence between different locations on the cortical surface.

2 Coregistration and Parameterization of Cortical Surfaces

In order to combine the recordings from different imaging modalities it is necessary to represent all data in a common coordinate system. This sounds like a trivial prerequisite but the different manufacturers of EEG and MEG

[1] An exception to this rule is the auditory brainstem response but in this case several thousand events have to be averaged in order to obtain a decent signal/noise ratio.

recording systems all have their specific preferences and a commonly accepted standard is nowhere in sight. Moreover, radiologists like to display the patient's left side on the right and vice versa, because this is the view they get when they look at them face to face.

2.1 Coregistration

The coordinate system we will use here is based on three landmarks on the human head that can easily be identified: the left and right pre-auricular points and the nasion. Our x-axis runs from the left to the right pre-auricular point, the y-axis runs through the nasion and has an intersection with the x-axis at a right angle. The z-axis is orthogonal to the xy-plane through the origin and points upwards to form a right-handed coordinate system. In a first step we will transform a structural MRI scan and the locations of EEG electrodes and MEG sensors into these coordinates.

Our starting point is a MRI volume of axial slices from a fullhead scan with a voxel size of $1\times1\times1\,mm^3$. From this scan the 3d-coordinates of three landmarks on the subject's head, i.e. the nasion and the left and right pre-auricular points, have to be found. This task is considerably simplified if vitamin E capsules have been attached to these points prior to the scan so that they can be easily identified as bright spots on a T_1 weighted scan due to the short spin-lattice relaxation time in its C-H_2 chains. The top row of Fig. 1 (left) shows examples for such landmarks in an axial, coronal and sagittal slice, respectively. In these images no two of the fiducials are located in the same slice. Our goal is to find a transformation matrix that defines a rotation and shift of the volume in 3d-space such that all three landmarks end up in the same axial slice which then defines the new xy-plane.

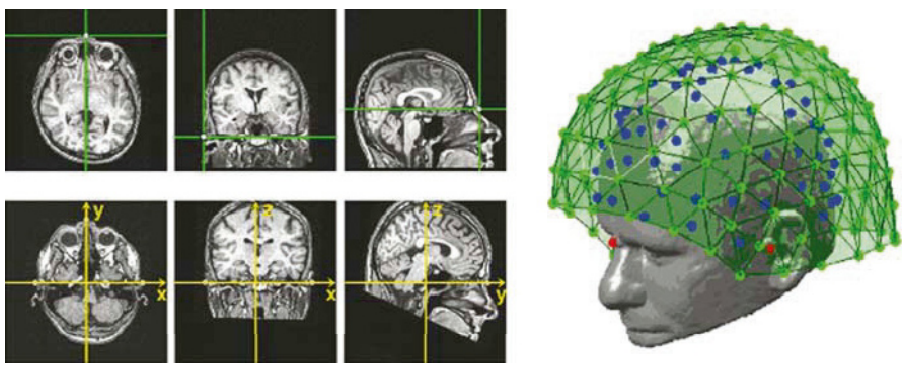


Fig. 1. Left top: Locations of fiducial points in a T_1 weighted fullhead MRI scan. Left bottom: The transformed volume. Note that all three landmarks are now located in the same axial slice, which represents the xy-plane. Right: The skin surface reconstructed from the MRI with fiducials (rod dots), EEG electrode locations (blue dots) and the surface defined by the locations of the MEG sensors (green)

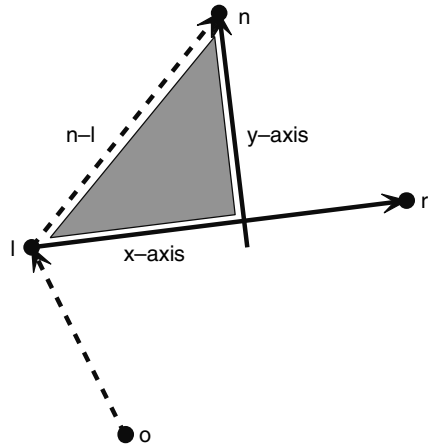


Fig. 2. Relation between the origin of the old coordinate system ($\boldsymbol{o}$), the landmarks ($\boldsymbol{l}$, $\boldsymbol{r}$ and $\boldsymbol{n}$) and the new x- and y-axis. The shaded triangle represents the vector relation (3)

Figure 2 shows a sketch of the relation between the landmarks ($\boldsymbol{r}$, $\boldsymbol{l}$, $\boldsymbol{n}$) and the old origin ($\boldsymbol{o}$). The new x-axis which runs from the left to the right pre-auricular is given by $\boldsymbol{l} + \alpha\,(\boldsymbol{r} - \boldsymbol{l})$. In the same way the xy-plane can be expressed as $\boldsymbol{l} + \gamma\,(\boldsymbol{r} - \boldsymbol{l}) + \beta\,(\boldsymbol{n} - \boldsymbol{l})$, where α, β and γ are arbitrary real numbers. The new y-axis is represented by the line in the xy-plane which is orthogonal to the x-axis and runs through the nasion ($\boldsymbol{n}$). All lines in the xy-plane perpendicular to the x-axis fulfill the condition

$$\{\gamma\,(\boldsymbol{r} - \boldsymbol{l}) + \beta\,(\boldsymbol{n} - \boldsymbol{l})\} \cdot \{\boldsymbol{r} - \boldsymbol{l}\} = 0 \tag{1}$$

which leads to a relation between γ and β, namely

$$\gamma = -\beta\,\frac{(\boldsymbol{n} - \boldsymbol{l}) \cdot (\boldsymbol{r} - \boldsymbol{l})}{|\boldsymbol{r} - \boldsymbol{l}|^2} \tag{2}$$

Using (2), the expression for the xy-plane and the abbreviations $\boldsymbol{r} - \boldsymbol{l} = \boldsymbol{r_l}$ and $\boldsymbol{n} - \boldsymbol{l} = \boldsymbol{n_l}$ the boundary of the shaded triangle in Fig. 2 can be expressed as

$$\alpha\,\boldsymbol{r_l} + \beta\,\{\boldsymbol{n_l} - \frac{\boldsymbol{n_l} \cdot \boldsymbol{r_l}}{|\boldsymbol{r_l}|^2}\,\boldsymbol{r_l}\} = \boldsymbol{n_l} \tag{3}$$

representing a set of three linear equations for the three vector components, and α and β can readily be calculated from any two of them. The origin of the new coordinate system is then given by $\boldsymbol{l} + \alpha\boldsymbol{r_l}$, for instance. The directions of the new axes can be written in the form

$$\begin{aligned} &\boldsymbol{r_l} && \text{for the } x\text{-axis,} \\ &\boldsymbol{n_l} - \tfrac{\boldsymbol{n_l} \cdot \boldsymbol{r_l}}{|\boldsymbol{r_l}|^2}\,\boldsymbol{r_l} && \text{for the } y\text{-axis, and} \\ &\boldsymbol{r_l} \times \{\boldsymbol{n_l} - \tfrac{\boldsymbol{n_l} \cdot \boldsymbol{r_l}}{|\boldsymbol{r_l}|^2}\,\boldsymbol{r_l}\} && \text{for the } z\text{-axis.} \end{aligned} \tag{4}$$

Once normalized, these vectors compose the rotation matrix $\boldsymbol{R}$ which defines the new coordinate systems after the origin has been shifted to its new location $\boldsymbol{l} + \alpha \boldsymbol{r_l}$.

In general, a rotation in three dimensional space can be described by consecutive rotations around three axes and parameterized by the Euler angles ϕ, θ and ψ. There are various ways to define these angles and the sequence of rotations. We will use here a form that first rotates the volume around the z-axis, then around the y-axis and finally around the x-axis. The corresponding transformation matrix $\boldsymbol{R}$ is the product of the matrices that describe the single rotations and reads explicitly

$$\begin{aligned} \boldsymbol{R} = \boldsymbol{R_x R_y R_z} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & -\sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ \sin\psi\sin\theta\cos\phi - \cos\psi\sin\phi & \sin\psi\sin\theta\sin\phi + \cos\psi\cos\phi & \sin\psi\cos\theta \\ \cos\psi\sin\theta\cos\phi + \sin\psi\sin\phi & \cos\psi\sin\theta\sin\phi - \sin\psi\cos\phi & \cos\psi\cos\theta \end{pmatrix} \end{aligned} \tag{5}$$

From the components R_{ij} of the rotation matrix on the right hand side of (5) we can determine the Euler angles as

$$\phi = \arctan\frac{R_{12}}{R_{11}} \qquad \theta = \arctan\frac{-R_{13}\sin\phi}{R_{12}} \qquad \psi = \arctan\frac{R_{23}}{R_{33}} \tag{6}$$

These angles together with the shift of the origin can be used in standard MRI software to create a new volume which has all three landmarks in the slice that now represents the xy-plane as shown in the bottom row of Fig. 1 (left).

Transforming the locations of electrodes or sensors into the the same coordinate system is straightforward. First the new origin is calculated according to (3) and the locations are shifted. Then a rotation with the normalized matrix (4) is performed which completes the coregistration procedure between the subject's head, and the electrodes and sensors as shown in Fig. 1 (right).

2.2 Parameterizing the Folded Brain Surfaces

The work of Dale, Sereno, Fischl and co-workers provided the scientific community with a powerful software tool known as *Freesurfer* (Dale and Sereno 1993, Dale et al. 1999, Fischl et al. 1999a). Freesurfer creates tessellations of cortical surfaces like the boundary between the gray and white matter or the boundary between the gray matter and the cerebral spinal fluid. The folded surfaces for each hemisphere can then be inflated by reducing the local curvature in an iterative process and eventually mapped onto the surface of a sphere. This transformation is unique and invertible, i.e. every point on the cortical surfaces corresponds to a single point on the sphere and vice versa

(a property mathematicians call an isomorphism). Moreover, the neighborhood relation between points, i.e. the topology, is preserved. Such a one to one mapping is possible because the spherical and the cortical surfaces are both singly connected and therefore topologically equivalent. For each cortical hemisphere *Freesurfer* creates files containing the vertex coordinates (of about 100,000 vertices) and the corresponding tessellation (of about 200,000 triangles) for the gray-white matter boundary, the boundary between the gray matter and the cerebral spinal fluid, the spherical surface, and surfaces from inflation steps as desired. The tessellation for all of these surfaces is the same and every point on the sphere has a unique partner on all the other surfaces. This allows for a parameterization which maps the 3-dimensional cartesian coordinates (x, y, z) of the vertices for each surface onto two angles, the latitude ψ and longitude φ on the sphere. In other words, for each pair of angles (ψ, φ) there exists a triple of coordinates (x, y, z), which corresponds to a point located on a cortical or inflated surface. An example for the color coded x-, y- and z-coordinates of the gray-white matter boundary is shown in Fig. 3 (top row) with φ running along the horizontal axis and ψ running vertically. The middle row shows a representation of the same quantities in polar coordinates where ψ runs in the radial and φ in the azimuthal direction. Contour lines for constant values of x, y and z (indicated in green) represent the surface in sagittal, coronal and axial slices, respectively, as shown in the bottom row.

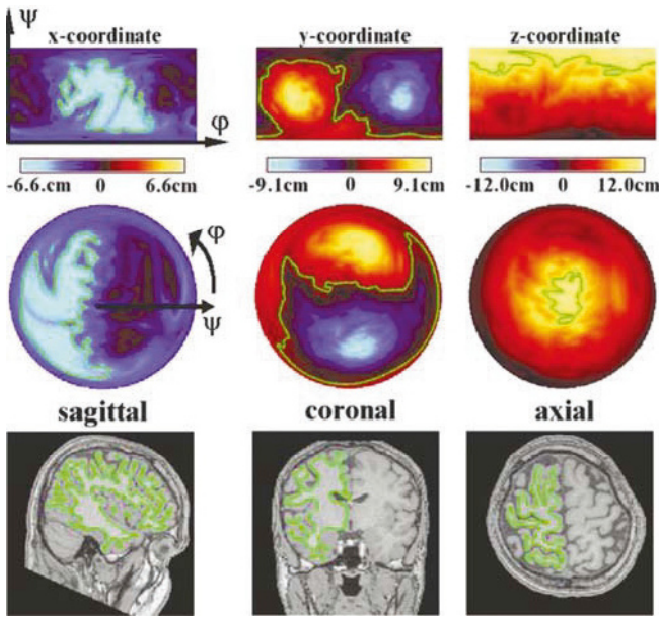


Fig. 3. Parameterization of the cortical surface. The x-, y- and z-coordinates of the folded three dimensional surface are expressed as continuous functions of the variables ψ and φ

Fig. 4. Inflation steps of the cortical surface into a sphere. Because of the topological equivalence between the folded surfaces and the inflated sphere, the spherical coordinate system can be mapped back onto the cortical surface

Such a parameterization in spherical coordinates also allows for defining a coordinate system on the folded surfaces as shown in Fig. 4. Such a coordinate system is useful for visualization purposes and for comparisons between different subjects (Fischl et al. 1999a, Fischl et al. 1999b).

3 The Dataset

Throughout this chapter we will use a dataset that simulates an MEG experiment where sources are active at two locations along the anterior and posterior wall of the central sulcus pointing into a direction perpendicular to the gray-white matter boundary as shown in Fig. 5.

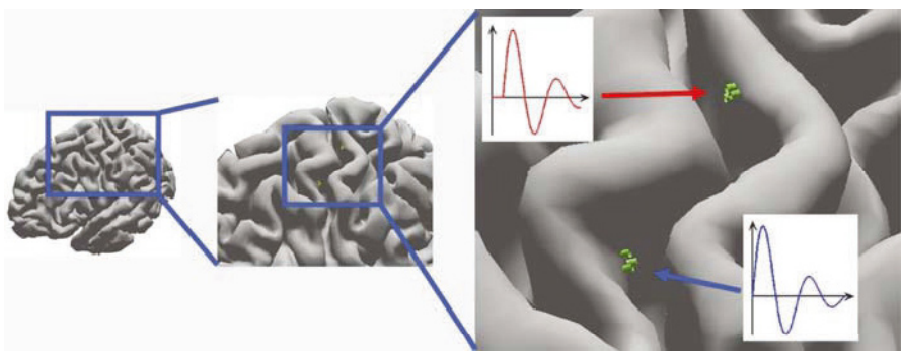


Fig. 5. Sources at two locations along the anterior and posterior wall of the central sulcus pointing into a direction perpendicular to the gray-white matter boundary. The red and blue curves in the inserts represent the time course of neural activity in these regions

We assume the time courses of activity at these locations to be damped oscillations plotted in the inserts, where the red curve is a time shifted version of the time series plotted in blue. We further assume that we measure the gradient of the radial component of the magnetic field at locations derived from a CTF Omega-151 magnetometer system (VSM MedTech Ltd., Port Coquitlam, BC, Canada), by calculating the superposition of magnetic fields form the individual six dipoles using the formula by (Sarvas 1987), which implicitly takes the contributions from the induced secondary currents in a spherical conducting volume into account.

$$\vec{B}(\vec{r}) = \frac{\mu_0}{4\pi F^2(\vec{r},\vec{r_q})}\{F(\vec{r},\vec{r_q})\,\vec{q}\times\vec{r_q} - [(\vec{q}\times\vec{r_q})\cdot\vec{r}]\,\nabla F(\vec{r},\vec{r_q})\} \tag{7}$$

where $\vec{r}$ denotes the location where the field is measured and $\vec{r_q}$ is the location of the dipolar current source with direction $\vec{q}$. The scalar function $F(\vec{r},\vec{r_q})$ and its gradient $\nabla F(\vec{r},\vec{r_q})$ are explicitly given by

$$\begin{aligned} F(\vec{r},\vec{r_q}) &= d\,(r\,d + r^2 - \vec{r_q}\cdot\vec{r}) \\ \nabla F(\vec{r},\vec{r_q}) &= \{\frac{d^2}{r} + \frac{\vec{d}\cdot\vec{r}}{d} + 2(d+r)\}\,\vec{r} - \{d + 2r + \frac{\vec{d}\cdot\vec{r}}{d}\}\,\vec{r_q} \end{aligned} \tag{8}$$

where the abbreviations $\vec{d} = \vec{r} - \vec{r_q}$, $d = |\vec{d}|$ and $r = |\vec{r}|$ have been used.

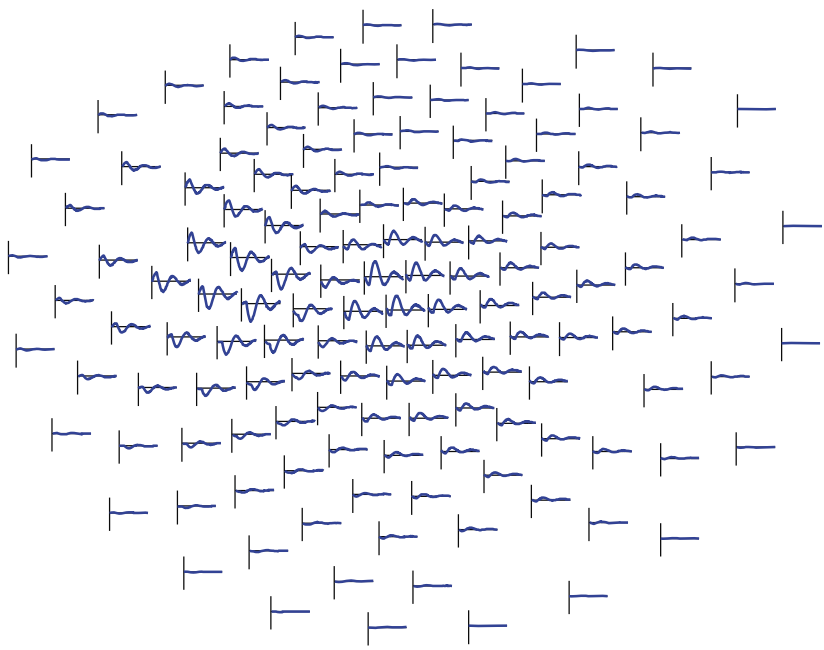


Fig. 6. Time series of the simulated data at 144 spatial locations, assuming that 7 of the 151 sensors were broken. Sensors are shown in polar projections with the nose on top

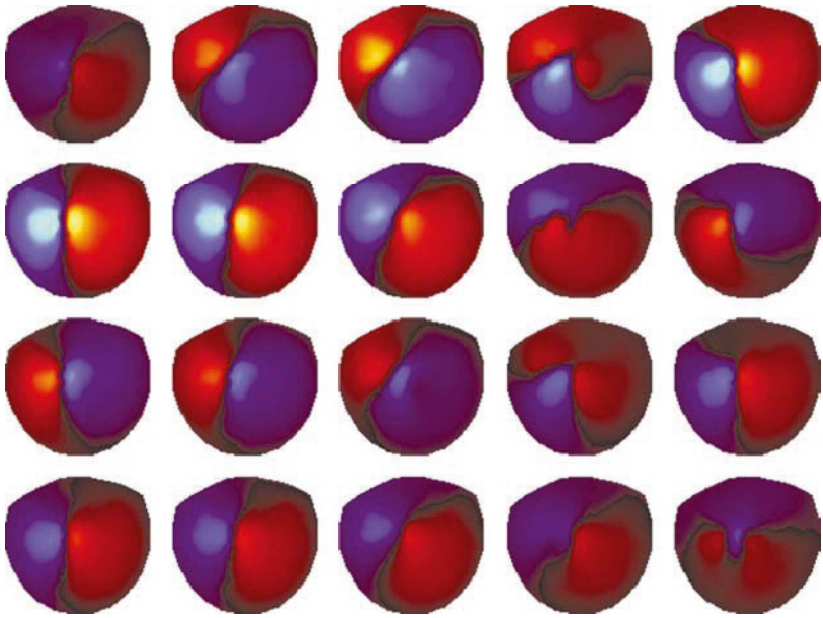


Fig. 7. Spatial patterns at 20 time points from top left to bottom right

The signal that is picked up by a sensing coil is given by the scalar product of the vector of the magnetic field and the normal vector of the coil area. A first order gradient is simulated by subtracting the signal measured at the outer coil from the signal picked up by the inner coil. In addition to the deterministic activity from the sources in the two active regions in the central sulcus two kinds of noise are added. First, to the sources themselves, and second to 500 locations on the cortical surface that are randomly chosen at each of the simulated 100 time steps. This way a total of 200 'trials' are created. Plots of the average time series across trials in a topological layout of the sensors as well as the spatial patterns at 20 points in time that would be measured in the given scenario are shown in Fig. 6 and Fig. 7, respectively.

4 Beamforming

4.1 The Basics

The basic idea behind beamforming in human brain research is to estimate the time course $q(t)$ of a current dipole at location $\vec{r}$ and direction $\vec{d}$ given measurements of one or more components of the magnetic field along a surface outside the head (MEG) or/and the electric potential on the scalp surface (EEG). We assume that we have recordings from M different sensor locations taken within

a time span T and lumped together to a vector $\vec{H}(t)$ representing the measured signal. The goal is to find a set of filter coefficients $\vec{w}$ such that the relation

$$q(t) = \vec{w} \cdot \vec{H}(t) \tag{9}$$

is fulfilled. The components of $\vec{w}$ are called the beamformer weights and are determined by minimizing the source power P over a time span T while keeping the projected signal constant $\vec{w} \cdot \vec{L} = 1$, where $\vec{L}$ is the forward solution or lead field (Moscher et al. 1999) of a unit source located at $\vec{r}$ with direction $\vec{d}$. The (average) source power can be expressed as

$$P = \frac{1}{T} \int_0^T q^2(t)\, dt = \frac{1}{T} \int_0^T \{\vec{w} \cdot \vec{H}(t)\}^2\, dt = \vec{w} \cdot \boldsymbol{C}\, \vec{w} \tag{10}$$

where $\boldsymbol{C}$ represents the correlation matrix $C_{ij} = \frac{1}{T} \int_0^T H_i(t)\, H_j(t)\, dt$.

The details on how to deal with such a minimization under constraints are described in appendix A where the beamforming vector $\vec{w}$ is explicitly derived as

$$\vec{w} = \frac{\boldsymbol{C}^{-1}\, \vec{L}}{\vec{L} \cdot \boldsymbol{C}^{-1}\, \vec{L}} \tag{11}$$

and where it is also shown that the source power P can be written in the form

$$P = \vec{w} \cdot \boldsymbol{C}\, \vec{w} = \{\vec{L} \cdot \boldsymbol{C}^{-1}\, \vec{L}\}^{-1} \tag{12}$$

As pointed out by Huang et al. (2004) the expression for the beamformer weights (11) and the global source power (12) are the same for all so-called linearly constrained minimum variance beamformers (LCMV) that exist in the literature. The differences between them are essentially threefold: First, in the way two quantities called the global and time dependent neural activity index are calculated. The global neural activation is a measure of activation in brain areas over a certain time span. The time dependent neural activity index is a time series representing local activations and is most meaningful at locations where the global index is large. The beamformers differ in the way these quantities are normalized, an issue originating from the inhomogeneous sensitivity of the beamformers with respect to depth and direction of the sources. An extreme case is MEG which is virtually blind to dipoles with a radial direction[2].

[2] It should be pointed out that this is not due to technological limitations and also has nothing to do with the fact that the CTF system we are using here measures mainly the radial component of the field. The reason is that a radial current in a spherical conductor does not produce any magnetic field outside the sphere which can be easily seen in the formula by Sarvas (7). As the direction of any radial current $\vec{q}$ is parallel to the position vector $\vec{r_q}$, their vector product is equal to $\vec{0}$, which implies $\vec{B}(\vec{r}) = \vec{0}$.

The second difference is how the beamformers treat low dimensional signals, i.e. signals with a low noise level like averages, where the correlation matrix $\boldsymbol{C}$ is singular and does not have an inverse. This problem can be dealt with by either subspace projection or regularization. The former reduces the dimensionality to the subspace defined by the eigenvectors of the correlation matrix which correspond to eigenvalues that are significantly bigger than zero. The latter is performed by adding a constant (representing uncorrelated noise) to the diagonal of the correlation matrix while keeping its original dimension. Both techniques, as discussed in detail in the following section, lead to matrices that are invertible. Explicit expressions for the global and time dependent neural activity index for the different types of beamformers found in the literature as distinguished by Huang et al. (2004) are given in table 1.

The third difference between the linearly constrained minimum variance beamformers concerns the way the underlaying brain volume is scanned. There are essentially three procedures to estimate the neural activity inside the brain. First there is vector beamforming: the volume is divided into voxels of a certain size (typically cubes with a side length of a few millimeters) and the global neural activity index for the center of each voxel is calculated three times for unit sources pointing into the x-, y- and z-direction. This way not only the activity is detected but also the direction of the source can be found. There are drawbacks using this method particularly with MEG, which is insensitive to radial currents, and therefore the estimate of a vector with a strong radial component is prone to large errors. This problem with radial components is avoided in the procedure introduced by Robinson and Vrba (1999) known in the literature as SAM (Synthetic Aperture Magnetometry). For this procedure within each voxel only tangential dipole directions are scanned and the direction showing the largest activity is used. It is this search in the tangential plane, which is different for every voxel as opposed to the fixed three components of the vector beamformers that has led some authors to call SAM a nonlinear procedure (Vrba and Robinson 2001). The third possibility to define locations and directions of potential sources, which we will use for demonstrating beamformers' performances later in this chapter, is to apply anatomical constraints where the currents are restricted to the cortical gray matter with a direction perpendicular to the gray-white matter boundary. This constraint is motivated by the fact that gray matter is organized in macrocolumns of pyramidal cells that act coherently to build assemblies of $10^4 - 10^5$ simultaneously active neurons which are necessary to produce a signal strong enough that it can be picked up by MEG sensors or EEG electrodes. The disadvantage here is the need for high-resolution structural MRIs from all subjects that are well coregistered with the sensor or electrode locations.

Table 1. Classification of the global neural activity index N_a and the time-dependent index $n_a(t)$ according to Huang et al. (2004). Here $\vec{L}$ represents the lead field, $\vec{H}(t)$ is the signal measured at the sensor locations, $\vec{w}$ the beamforming filter, $\boldsymbol{C}$ is the covariance matrix of the signal at the times of interest and $\boldsymbol{\Sigma}$ the covariance matrix of the noise. $\boldsymbol{\Sigma}$ is either estimated from baseline data or chosen as a constant times the identity matrix

Type	Global activity index	Time dependent activity index	Ref.
I	$N_a = \frac{\{\vec{L}\cdot\boldsymbol{C}^{-1}\,\vec{L}\}^{-1}}{\{\vec{L}\cdot\boldsymbol{\Sigma}^{-1}\,\vec{L}\}^{-1}} = \frac{\vec{L}\cdot\boldsymbol{\Sigma}^{-1}\,\vec{L}}{\vec{L}\cdot\boldsymbol{C}^{-1}\,\vec{L}}$	$n_a(t) = \frac{q^2(t)}{\{\vec{L}\cdot\boldsymbol{\Sigma}^{-1}\,\vec{L}\}^{-1}} = \frac{\{\vec{w}\cdot\vec{H}(t)\}^2}{\{\vec{L}\cdot\boldsymbol{C}^{-1}\,\vec{L}\}^{-1}}$	(Van Veen et al. 1997)
II	$N_a = \frac{\vec{w}\cdot\boldsymbol{C}\,\vec{w}}{\lvert\vec{w}\rvert^2} = \frac{\vec{L}\cdot\boldsymbol{C}^{-1}\,\vec{L}}{\vec{L}\cdot\boldsymbol{C}^{-2}\,\vec{L}}$	$n_a(t) = \frac{\{\vec{w}\cdot\vec{H}(t)\}^2}{\lvert\vec{w}\rvert^2} = \frac{\{\vec{L}\cdot\boldsymbol{C}^{-1}\,\vec{H}(t)\}^2}{\vec{L}\cdot\boldsymbol{C}^{-2}\,\vec{L}}$	(Borgiotti and Kaplan 1979), (Sekihara et al. 2001)
III	$N_a = \frac{\vec{w}\cdot\boldsymbol{C}\,\vec{w}}{\vec{w}\cdot\boldsymbol{\Sigma}\,\vec{w}} = \frac{\vec{L}\cdot\boldsymbol{C}^{-1}\,\vec{L}}{\vec{L}\cdot\boldsymbol{C}^{-1}\,\boldsymbol{\Sigma}\,\boldsymbol{C}^{-1}\,\vec{L}}$	$n_a(t) = \frac{\{\vec{w}\cdot\vec{H}(t)\}^2}{\vec{w}\cdot\boldsymbol{\Sigma}\,\vec{w}} = \frac{\{\vec{L}\cdot\boldsymbol{C}^{-1}\,\vec{H}(t)\}^2}{\vec{L}\cdot\boldsymbol{C}^{-1}\,\boldsymbol{\Sigma}\,\boldsymbol{C}^{-1}\,\vec{L}}$	(Robinson and Vrba 1999)
IV	$N_a = \frac{\{\vec{L}\cdot\boldsymbol{C}^{-n}\,\vec{L}\}^{-1}}{\{\vec{L}\cdot\boldsymbol{\Sigma}^{-n}\,\vec{L}\}^{-1}} = \frac{\vec{L}\cdot\boldsymbol{\Sigma}^{-n}\,\vec{L}}{\vec{L}\cdot\boldsymbol{C}^{-n}\,\vec{L}}$	$n_a(t) = \frac{\vec{w}\cdot\{\vec{H}(t)\cdot\vec{H}(t)\}^n\vec{w}}{\vec{L}\cdot\boldsymbol{\Sigma}^{-n}\,\vec{L}} = \frac{\lvert\vec{H}(t)\rvert^{2n}\lvert\vec{w}\rvert^2}{\vec{L}\cdot\boldsymbol{\Sigma}^{-n}\,\vec{L}}$	(Huang et al. 2004)

4.2 Beamformer Performance: Regularization and Subspace Expansion

Independent of which one of the beamformer types is used it is always necessary to obtain the inverse of the covariance matrix[3] $\boldsymbol{C}$. An inverse may not exist or its computation can be numerically unstable if $\boldsymbol{C}$ is close to singular which happens if the signal space has a dimension smaller than the number of sensors. In a covariance matrix calculated from averaged signals, for instance, there are typically only a few eigenvalues that are significantly different from zero and the matrix is at least numerically singular. As mentioned above there exist two ways to work around this problem called regularization and subspace projection.

In regularization a constant σ is added to the diagonal of the covariance matrix in order to create a non-singular matrix. The effects of regularization are shown in Fig. 8 for both the global and time dependent neural activity index calculated from the dataset described previously using a beamformer of type III. In contrast to Huang et al.'s version given in the table, we will not use the square in the numerator for the time dependent activity index in order to preserve direction of current flow. The global activity index is color-coded from red to yellow and thresholded at the same level for all four plots at 30% of its maximum. The inserts show the time series (solid) representing the time dependent activity index on top left in red for the active region in the posterior wall of the central sulcus and on bottom right in blue for the source in the anterior wall. The dashed lines in the inserts represent the ideal response.

In the top left plot a regularization parameter of $\sigma = 5\times10^{-4}$ is used. The time series for the time dependent activity index are in good agreement with the ideal shape, but the global index shows activity in many regions where no source is present. The determinant of the covariance matrix in this case is too small so that numerical errors prevent the calculation of a correct inverse. In the plot on the top right the regularization is $\sigma = 5\times10^{-3}$ and the matrix is well regularized. The time series overlap well with the ideal curves and the global activity index is focused and localized at the correct spots. In the lower left σ is 0.05. The spots of activity become slightly blurred and the time series start to show a phase shift with respect to the ideal functions indicating a superposition of the two original curves. These effects are even more pronounced on the bottom right with $\sigma = 0.5$, where over-regularization is obvious: the activity is blurred over wide regions and the time dependent index is clearly shifted, i.e. we obtain superpositions of the two time series. In short, if the regularization parameter is too small numerical instabilities distort the beamformer performance, if σ is too big the beamformer becomes

[3] The covariance matrix is the correlation matrix scaled such that the sum of all its eigenvalues is equal to 1. As we give numerical values for the regularization parameters this is a more appropriate representation than the correlation matrix, since the values in the covariance matrix do not depend on the units the magnetic field is measured in.

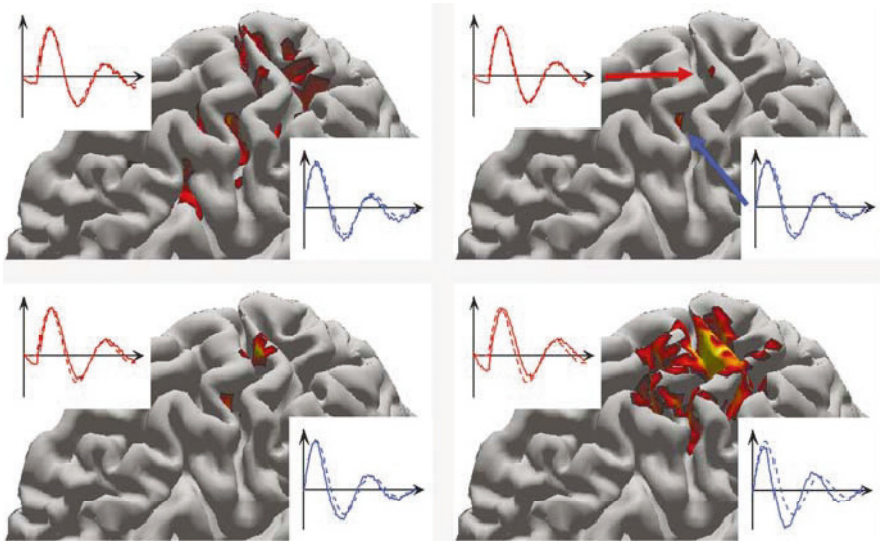


Fig. 8. Global neural activity index on the cortical surface for four different regularizations calculated from a beamformer of type III. From left top to right bottom the regularization parameter σ is increased from $5{\times}10^{-4}$ to 0.5 by factors of 10. On the top left the regularization parameter is too small and numerical instabilities distort the beamformer performance. In the bottom row σ is too big, the beamformer becomes less focussed or even blurred and activity from other regions leaks into the locations of interest. Inserts show the time dependent index reconstructed at the regions of maximum activity (solid) and the ideal curve (dashed). Especially on the bottom right a phase shift between these curves can be seen that originates from leakage of activity between the two regions due to a lack of focus

less focussed or even blurred and activity from other regions leaks into the locations of interest. Unfortunately, σ is not known upfront and needs to be determined by trial and error.

The subspace projection is done by using the subspace defined by the eigenvectors of the covariance matrix whose corresponding eigenvalues are sufficiently different from zero. Subspace expansion of the type III beamformer is shown in appendix A and the beamformer weights and the neural activity indices with the proper normalizations turn out to be

$$\vec{w} = \sum_{k=1}^{N} \frac{\{\vec{L}\cdot\vec{v}^{(k)}\}^2}{\lambda^{(k)}} \{\sum_{n=1}^{N} \frac{\{\vec{L}\cdot\vec{v}^{(n)}\}^2}{\lambda^{(n)}}\}^{-1} \vec{v}^{(k)}$$

$$N_a = \frac{\sum_{n=1}^{N}\{\vec{L}\cdot\vec{v}^{(n)}\}^2}{\sum_{n=1}^{N}\frac{\{\vec{L}\cdot\vec{v}^{(n)}\}^2}{\lambda^{(n)}}} \tag{13}$$

$$n_a(t) = \frac{\vec{w}\cdot\vec{H}(t)}{|\vec{w}|}$$

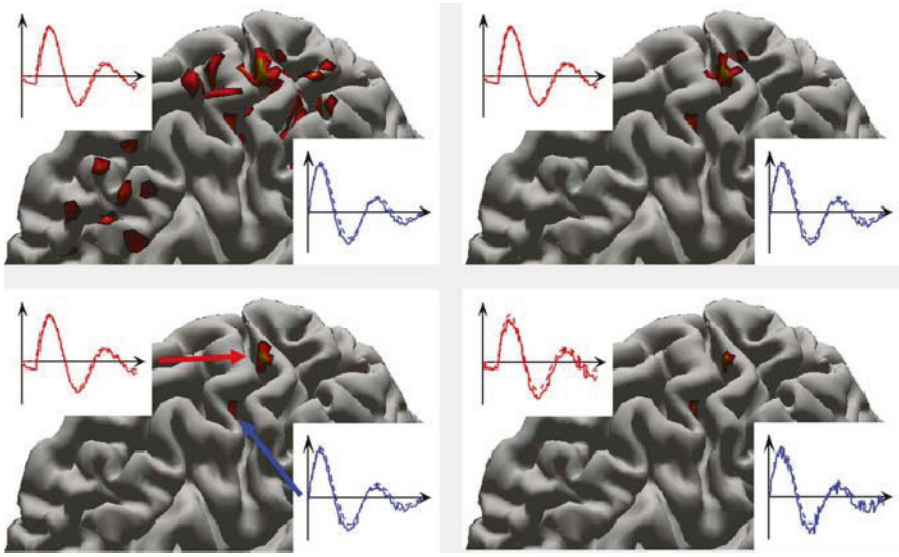


Fig. 9. Global neural activity index on the cortical surface for four different subspace dimensions calculated from a beamformer using (13). From top left to bottom right the subspace dimensions are 4, 8, 12 and 50, respectively. The global activity index is blurred if not enough dimensions are taken into account and gets more focused as this number increases. On the other hand, if too many dimensions are used the time dependent index becomes contaminated by noise

where $\lambda^{(n)}$ and $\vec{v}^{(n)}$ are the eigenvalues and eigenvectors of the covariance matrix, respectively.

Figure 9 shows the effects of subspace projection when taking different numbers of eigenvalues and eigenvectors into account. Specifically, starting at the top left to bottom right these numbers are 4, 8, 12 and 50, respectively. The global activity index is blurred if not enough dimensions are taken into account and gets more focused as this number increases. On the other hand, if too many dimensions are used the time dependent index becomes contaminated by noise.

5 Coherence and Granger Causality

Statistical measures have been used for a long time to study synchronization and desynchronization in higher frequency bands in datasets from EEG/MEG recordings (see e.g. (Pfurtscheller and Lopes da Silva 1999)) and measurements of the local field potential from implanted electrodes (Brovelli et al. 2004). Here we will establish coherence and Granger causality obtained from the time dependent neural activity index of the two active regions on the cortical surface as measures that allow for detecting connections between different brain regions or functional brain connectivity in general. We will briefly

describe how these quantities are defined and calculated, and refer the reader to the chapter by Kaminski (this volume) for more detailed discussions on related topics.

5.1 Coherence

Coherence is calculated starting from two time series $x(t)$ and $y(t)$ and obtaining their Fourier transforms

$$X(\omega) = \int_0^T x(t)\, e^{j\omega t}\, dt \qquad \text{and} \qquad Y(\omega) = \int_0^T y(t)\, e^{i\omega t}\, dt \tag{14}$$

which leads to their spectral density functions

$$\begin{aligned} S_{xx}(\omega) &= \frac{1}{T}\, \big|X(\omega)\big|^2 \qquad S_{yy}(\omega) = \frac{1}{T}\, \big|Y(\omega)\big|^2 \\ S_{xy}(\omega) &= \frac{1}{T}\, X^*(\omega)\, Y(\omega) = S_{yx}(\omega) = \frac{1}{T}\, Y^*(\omega)\, X(\omega) \end{aligned} \tag{15}$$

From these functions the so-called squared coherence spectrum can be calculated as

$$C_{xy}(\omega) = \frac{\big|S_{xy}(\omega)\big|^2}{S_{xx}(\omega)\, S_{yy}(\omega)} \tag{16}$$

and we can define a global coherence as the integral of $C_{xy}(\omega)$ across all frequencies.

5.2 Granger Causality

From the coherence between two signals we can conclude whether they have an influence on each other. We cannot identify the direction, i.e. whether x drives y or y drives x, or whether there is mutual feedback between the two. A quantity which allows for such a distinction is a measure called Granger causality, where the word 'causality' in its name is a rather unlucky choice. Given two time series $x(t)$ and $y(t)$ we can never determine by any kind of analysis alone whether there is a causal relation between them, let alone who causes who. What Granger causality does allow to determine is whether the knowledge of one of the time series enables us to make a better prediction about the other one than would have been possible without that knowledge.

Granger causality is usually calculated from multivariate autoregressive models. To this end we assume that we have two time series x_t and y_t sampled at discrete[4] times t. Now we can set up autoregressive models for x_t andy_t,

[4] We assume discrete time steps for simplicity, for continuous time series an embedding has to be determined, see Chen et al. (2004) for details.

i.e. express their values at times t in terms of their past

$$x_t = \sum_{k=1}^{p} a_k \, x_{t-k} + \epsilon_t^{(x)} \qquad y_t = \sum_{k=1}^{p} c_k \, y_{t-k} + \epsilon_t^{(y)} \tag{17}$$

where we take p previous values of the time series into account to obtain a prediction of x_t and y_t. The equations in (17) are models of order p and the coefficients a_k and c_k have to be determined across all sequences in the time series. The difference between the predicted and the actual values for each x_t and y_t is given by the errors $\epsilon_t^{(x)}$ and $\epsilon_t^{(y)}$, respectively. Similarly we can set up multivariate autoregressive models where x_t does not only depend on its own past but also on previous values of y_t and vice versa

$$x_t = \sum_{k=1}^{p} a_k \, x_{t-k} + b_k \, y_{t-k} + \epsilon_t^{(x|y)} \qquad y_t = \sum_{k=1}^{p} c_k \, y_{t-k} + d_k \, x_{t-k} + \epsilon_t^{(y|x)} \tag{18}$$

We now say that y has a Granger causal influence on x if including y in the model for x on average improves the model, i.e. decreases the error

$$\frac{< \epsilon_t^{(x|y)} >}{< \epsilon_t^{(x)} >} < 1 \tag{19}$$

where $< \ldots >$ denotes the expectation value.

There are different ways to quantify the Granger causal influence of one time series onto another. One straightforward definition is evidently by means of the errors

$$g_{y \to x} = 1 - \frac{< \epsilon_t^{(x|y)} >}{< \epsilon_t^{(x)} >} \tag{20}$$

which is a number between zero and one with $g_{y \to x} = 0$ indicating that there is no influence.

5.3 Application to the Dataset

We applied the coherence measure and Granger causality to data sets that were constructed in the following way. A beamformer was separately applied to the 200 trials of simulated MEG data. Then the time series from each trial at the two active locations were extracted, the average time series was subtracted from each of the realizations and concatenated for each of the two regions. The time series corresponding to the red curve in Fig. 5, we shall call $x(t)$, is a shifted version of the blue curve $y(t)$, which can be interpreted as originating from a one-directional coupling from the lower to the upper brain region with the finite distance between them leading to a time delay. We can vary the coupling strength between the two regions by gradually varying the second time series from random noise to $y(t)$. We therefore define a time series $y_c(t)$ as

$$y_c(t) = c \, y(t) + (1 - c) \, \xi(t) \qquad \text{with} \qquad 0 \leq c \leq 1 \tag{21}$$

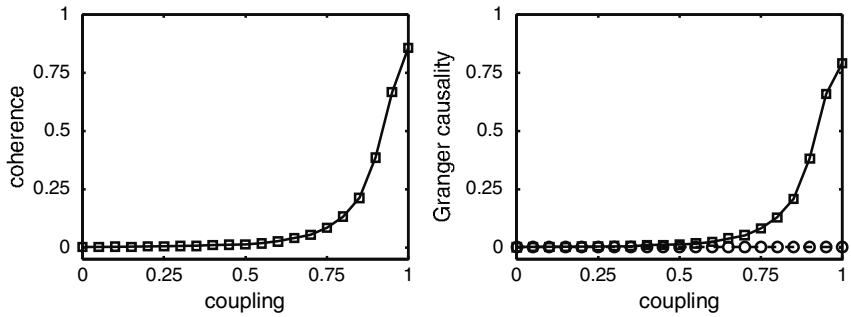


Fig. 10. Coherence (left) and Granger causality (right) as a function of the coupling strength between the two active regions. On the right, squares indicate a feed from the lower into the upper region whereas circles stand for a feed into the opposite direction. We can conclude that there is a one-directional connection between the two regions as the dashed line with circles stays close to zero for all couplings

where $\xi(t)$ represents gaussian white noise. Plots of coherence and Granger causality as functions of coupling strength c are shown in Fig. 10. In the left plot the coherence between the two time series $x(t)$ and $y_c(t)$ increases with increasing coupling strength c. On the right, the line with squares represents the Granger causal influence $g_{y_c \to x}$ of $y_c(t)$ on $x(t)$ which also increases with coupling c, whereas $g_{x \to y_c}$ represented by the line with circles stays close to zero independent of c. We can therefore conclude that there is a one-directional connection between the two brain regions where the lower region feeds into the upper but not vice versa.

6 Conclusions and Outlook

Noninvasive recording of human brain activity has undergone a major change during the last decades with the development of high-density electrode and sensor grids in EEG and MEG, the introduction of MRI and functional MRI, and the explosion in computational power and storage capacity, representing one side of the coin: advances in technology. On the other side of this coin is written: "What do we get from all this stuff and how do we make sense of it?" How do we extract relevant information from a humongous tangled mass of data. Even though "relevant information" is a subjective matter that strongly depends on the specific questions one is interested in, relying on a single technology is in most cases not a good idea. In this chapter we have tried to show how different imaging technologies can be used to implement anatomical constraints on potential sources whose time series can then be identified from electrophysiological recordings and, moreover, give us insight regarding the existence and nature of a connection between two brain areas. There are other independent methods like Diffusion Tensor Imaging (Westin

et al. 2002) or Structural Equation Modeling of functional MRI data (McIntosh and Gonzalez-Lima 1994) that can be used to probe connectivity, and only if different methods converge onto the same results can we have confidence that we have found something that really exists. In any case, there is no doubt that beamforming can become and probably will become one of the pillars that a future understanding of brain connectivity will be built on.

Acknowledgements

We thank Mingzhou Ding and Yonghong Chen for providing their Matlab code that was used to calculate coherence and Granger causality. Research supported by NINDS grant 1R01 NS48229.

A Minimizing the Source Power under Constraints

Here we show explicitly how the beamformer coefficients $\vec{w}$ and the source power P as given in (12) can be derived. Our starting point is a signal $\vec{H}(t)$ from EEG or MEG recordings where each of the M component $H^i(t)$ represents a time series from a single channel. Even though in practice these time series will be sampled at a finite rate we will treat time here as a continuous variable because the notation is less confusing that way. It is our goal to find a set of coefficients $\vec{w_\theta}$ such that the sensor array becomes most sensitive to a current source at a certain location (x, y, z) and a certain direction (ϑ, φ) which we will abbreviate by $\vec{\theta} = \vec{\theta}(x, y, z, \vartheta, \varphi)$. Applying the beamformer $\vec{w_\theta}$ to the signal $\vec{H}(t)$ we can write the global source power originating at $\vec{\theta}$ in the form

$$\begin{aligned} P_\theta &= \tfrac{1}{T} \textstyle\int_0^T dt \, \{\vec{w_\theta} \cdot \vec{H}(t)\}^2 \\ &= \tfrac{1}{T} \textstyle\int_0^T dt \, \sum_{i=1}^M \{w_\theta^i \, H^i(t)\}^2 \\ &= \sum_{i=1}^M \sum_{j=1}^M w_\theta^i \, w_\theta^j \underbrace{\frac{1}{T} \int_0^T H^i(t) \, H^j(t) \, dt}_{C_{ij}} \end{aligned} \tag{22}$$

where C_{ij} represents the correlation matrix. Using $\boldsymbol{C}$, the global source power originating from $\vec{\theta}$ can now be written in the compact form

$$P_\theta = \sum_{i=1}^M \sum_{j=1}^M C_{ij} \, w_\theta^i \, w_\theta^j = \vec{w_\theta} \cdot \boldsymbol{C} \, \vec{w_\theta} \tag{23}$$

The goal of finding a vector $\vec{w_\theta}$ such that the sensor array becomes most sensitive to a current source at $\vec{\theta}$ is achieved by minimizing the global power

from all locations and directions while keeping the signal originating at $\vec{\theta}$ constant. In other words

$$P_\theta = \vec{w_\theta} \cdot \boldsymbol{C}\, \vec{w_\theta} = \text{Min} \quad \text{with the constraint} \quad \vec{w_\theta} \cdot \vec{L_\theta} = 1 \tag{24}$$

where $\vec{L_\theta}$ is the forward solution or lead field from a unit current source at $\vec{\theta}$. The constraint in (24) means that the components of $\vec{w_\theta}$ cannot be varied independently in order to find the minimum but only combinations for which the constraint is fulfilled are allowed. Without the constraint an obvious solution for the minimum would be $\vec{w_\theta} \equiv \vec{0}$, which is certainly not what we are looking for, and it also violates the constraint.

The problem of finding the minimum under constraints of the form (24) is well known in physics and engineering, and can be solved using the method of Lagrange multipliers. The idea is to rewrite the constraint in the form

$$\vec{w_\theta} \cdot \vec{L_\theta} - 1 = 0 \tag{25}$$

where now the expression on the left hand side of (25) vanishes. It still vanishes if we multiply it by a constant λ, and it does not change anything if we add it to the global power we want to minimize, because we are essentially adding zero

$$P_\theta = \sum_{i=1}^{M} \sum_{j=1}^{M} C_{ij}\, w_{\vec{\theta}}^i w_{\vec{\theta}}^j + \lambda \{ \sum_{i=1}^{M} w_{\vec{\theta}}^i L_{\vec{\theta}}^i - 1 \} = \text{Min} \tag{26}$$

But now we can vary the components of $\vec{w_\theta}$ independently and we find the minimum by taking the derivative of P_θ with respect to w_θ^k and solve

$$\frac{\partial P_{\boldsymbol{\theta}}}{\partial w_\theta^k} = 2 \sum_{i=1}^{M} C_{ik}\, w_\theta^i + \lambda\, L_\theta^k = 0 \tag{27}$$

for $\vec{w_\theta}$. If we rewrite (27) in matrix form

$$2\, \boldsymbol{C}\, \vec{w_\theta} = -\lambda\, \vec{L_\theta} \tag{28}$$

we find immediately

$$\vec{w_\theta} = -\frac{\lambda}{2}\, \boldsymbol{C}^{-1}\, \vec{L_\theta} \tag{29}$$

This solution still depends on the unknown Lagrange parameter λ, which can, however, be determined by inserting (29) into (25) leading to

$$\{ -\frac{\lambda}{2}\, \boldsymbol{C}^{-1}\, \vec{L_\theta} \}\, \vec{L_\theta} = 1 \quad \text{or} \quad \lambda = -2\, \{ \vec{L_\theta} \cdot \boldsymbol{C}^{-1}\, \vec{L_\theta} \}^{-1} \tag{30}$$

By inserting (30) into (29) we finally obtain for the beamformer coefficients

$$\vec{w_\theta} = \frac{\boldsymbol{C}^{-1}\,\vec{L_\theta}}{\vec{L_\theta}\cdot\boldsymbol{C}^{-1}\,\vec{L_\theta}} \tag{31}$$

and for the global source power

$$P_\theta = \vec{w_\theta}\cdot\boldsymbol{C}\,\vec{w_\theta} = \{\vec{L_\theta}\cdot\boldsymbol{C}^{-1}\,\vec{L_\theta}\}^{-1} \tag{32}$$

As mentioned previously, if the signal space is low-dimensional as for averaged signals for instance, the correlation matrix $\boldsymbol{C}$ is singular and does not have an inverse. In this case a projection onto the relevant subspace can be achieved by expanding the beamformer vector $\vec{w}$ and the lead field $\vec{L}$ into the eigenvectors $\vec{v}^{(k)}$ of the correlation matrix that correspond to the N eigenvalues $\lambda^{(k)}$ that are significantly bigger than zero[5]

$$\vec{w} = \sum_{n=1}^{N}\omega_n\vec{v}^{(n)} \qquad \text{and} \qquad \vec{L} = \sum_{n=1}^{N}\zeta_n\vec{v}^{(n)} \tag{33}$$

Inserting (33) into (31) and (32) we obtain after straightforward calculations

$$\omega_k = \frac{\zeta_k}{\lambda^{(k)}}\{\sum_{n=1}^{N}\frac{\zeta_n^2}{\lambda^{(n)}}\}^{-1} \qquad \text{and} \qquad P = \{\sum_{n=1}^{N}\frac{\zeta_n^2}{\lambda^{(n)}}\}^{-1} \tag{34}$$

References

Abeles M (1991) *Corticonics: Neural Circuits of the Cerebral Cortex*, Cambridge University Press, Cambridge

Borgiotti GV, Kaplan LJ (1979) Superresolution of Uncorrelated Interference Sources by Using Adaptive Array Techniques. IEEE Trans. Antennas Propagat. 27:842–845

Braitenberg V, Schüz A (1991) *Cortex: Statistics and Geometry of Neural Connectivity*, Springer, Berlin

Brovelli A, Ding M, Ledberg A, Chen Y, Nakamura R, Bressler SL (2004) Beta oscillations in a large-scale sensorimotor cortical network: Directional influence revealed by Granger causality. PNAS 101:9849–9854.

Chen Y, Rangarajan G, Feng J, Ding M (2004) Ananlyzing Multiple Nonlinear Time Series with Extended Granger Causality. Phys. Lett. A 324:26–35

Dale AM, Fischl B, Sereno MI (1999) Cortical Surface-Based Analysis I: Segmentation and Surface Reconstruction. Neuroimage 9:179–194.

Dale AM, Sereno MI (1993) Improved Location of Cortical Activity by Combining EEG and MEG with MRI Cortical Surface Reconstruction: A Linear Approach. Journal of Congnitive Neuroscience 5:162–176.

Fischl B, Sereno MI, Dale AM, (1999a) Cortical Surface-Based Analysis II: Inflation Flattening and a Surface-Based Coordinate System. Neuroimage 9:194–207.

[5] As the correlation matrix is real, symmetric and positive semi-definite all eigenvalues are real and non-negative, and the eigenvectors are mutually orthogonal.

Fischl B, Sereno MI, Tootell RBH, Dale AM (1999b) High-Resolution Intersubject Averaging and a Coordinate System for the Cortical Surface. Human Brain Mapping 8:272–284.

Frost III OL, (1972) An Algorithm for Linearly Adaptive Array Processing. Proc. IEEE 60:926–935

Fuchs A (2002) Combining Technologies: The Use of Brain Surfaces. In: *Biomag 2002, Proc. 13th Int. Conf. on Biomagnetism*, Nowak H, Haueisen J, Geißler F, Huonker R, eds., VDE Verlag Berlin pp. 878–880

Hillebrand A, Barnes GR (2003) The use of anatomical constraints with MEG beamformers. Neuroimage 20:2302–2312

Huang M-X, Shih JJ, Lee RR, Harrington DL, Thoma RJ, Weisend MP, Hanion F, Paulson KM, Li T, Martin K, Miller GA, Canive JM (2004) Commonalities and Differences Among Vectorized Beamformers in Electromagnetic Source Imaging. Brain Topography 16:139–158

McIntosh AR, Gonzalez-Lima F (1994) Structural equation modeling and its application to network analysis in functional brain imaging. Human Brain Mapping 2:2–22.

Moscher JC, Leahy RM, Lewis PS (1999) EEG and MEG Forward Solutions for Inverse Models. IEEE Trans. Biomed. Eng. 46:245–259

Nunez PL (1981) *Electric Fields of the Brain*, Oxford University Press, Oxford

Pfurtscheller G, Lopes da Silva FH (1999) Event-related EEG/MEG synchronization and desynchronization: beasic principles. Clinical Neurophysiology 110: 1842–1857

Robinson SE, Vrba J (1999) Functinal neuroimaging by synthetic aperture magnetometry. In: *Recent advances in biomagnetism*, Yoshimito T, Kotani M, Kuriki S, Nagasato N, eds., Tohoku University Press, Sedai, Japan pp. 302–305.

Sarvas J (1987) Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. Phys. Med. Biol. 32:11–22.

Sekihara K (1996) Generalized Wiener Estimation of Three-Dimensional Current Distribution from Biomagnetc Measurements. IEEE Trans. Biomed. Eng. 43:281–291

Sekihara K, Nagarajan SS, Poeppel D, Maranntz A, Miyashita Y (2001) Reconstructing Spatio-Temporal Activities of Neural Sources Using an MEG Vector Beamformer Technique. IEEE Trans. Biomed. Eng. 48:760–771

Van Veen BD, Buckley KM (1988) Beamforming: A Versatile Approach to Spatial Filtering. IEEE ASSP Mag. 5:4–24

Van Veen BD, van Drongelen W, Yuchtman M, Suzuki A (1997) Localization of Brain Electrical Activity via Linearly Constraint Minimum Variance Spatial Filtering. IEEE Trans. Biomed. Eng. 44:867–880

Vrba J, Robinson SE (2001) Differences between synthetic aperture magnetometry (SAM) and linear beamformers. In: *Biomag 2000, Proc. 12th Int. Conf. on Biomagnetism*, Nenonen J, Ilmoniemi RJ, Katila T, eds., Helsinki University of Technology, Espoo, Finland pp. 681–684.

Westin CF, Maier SE, Mamata H, Nabavi A, Jolesz FA, Kikinis R (2002) Processing and visualization for diffusion tensor MRI. Medical Image Analysis 6:93–108

Frequency-Dependent Functional Connectivity Analysis of fMRI Data in Fourier and Wavelet Domains

R Salvador[1,2], S Achard[2] and ET Bullmore[2]

[1] Sant Joan de Déu - SMM, Fundació Sant Joan de Déu. 08830, Sant Boi de Llobregat (Barcelona), Spain
[2] Brain Mapping Unit and Wolfson Brain Imaging Centre, University of Cambridge, Departments of Psychiatry and Clinical Neurosciences, Addenbrooke's Hospital, Cambridge CB2 2QQ, UK

The analysis of physiological relationships between multiple brain regions has been spurred by the emergence of functional magnetic resonance imaging (fMRI), which provides high spatial resolution of dynamic processes in the human brain with a time resolution in the order of seconds. Many different conceptual approaches and algorithmic solutions have been proposed for connectivity analysis in the last 10 years or so (see some examples in Bullmore et al 1996, McIntosh 1999, Horwitz 2003, Ramnani et al 2004). These can be broadly sub-divided into analyses of functional connectivity – usually defined as a statistical association between spatially remote time series – or effective connectivity – the causal influence that one time series exerts over another (Espinosa & Gerstein 1988, Friston et al. 1997). While functional connectivity between brain regions is frequently depicted through undirected graphs, effective connectivity networks are more naturally portrayed by directed graphs.

In addition to this well-rehearsed distinction, we can also categorise available methods according to the mathematical domain in which they are implemented. While the great majority of methods for both functional and effective connectivity analysis of fMRI data have been implemented in the time domain, considerably fewer methods to date have been implemented in the Fourier domain (Cordes et al. 2001, Sun et al. 2004, Salvador et al. 2005a, Yamashita et al. 2005); and the development of approaches in the wavelet domain has been very recent (Achard et al. 2006). In comparable analyses of brain functional networks based on electromagnetic (EEG or MEG) data, the adoption of tools in the Fourier and wavelet domains has been more widespread (see reviews by Samar et al. 1995 and Koenig et al. 2005).

The main motivation for further consideration of Fourier and wavelet methods for functional connectivity analysis of fMRI data is that associations between brain regions may not be equally subtended by all frequencies; rather, some frequency bands may be of special importance in mediating functional

connectivity. There is, for example, abundant prior evidence that functional connectivity measured with subjects lying quietly in the scanner at "rest" is subtended predominantly by very low frequencies, <0.1 Hz, for many pairs of connected regions (Biswal et al. 1995, Lowe et al. 2000, Cordes et al. 2001, Robouts et al. 2003). This presumably reflects at a bivariate or multivariate level of analysis the well-replicated but still incompletely understood phenomenon of low frequency endogenous oscillations in resting fMRI (Maxim et al. 2005, Salvador et al. 2005b), optical imaging (Mayhew et al. 1996) and EEG time series (Leopold & Logothetis 2003). Functional connectivity analysis using tools that naturally support the frequency decomposition of physiological associations between regions may therefore be of interest in "denoising" the analysis (by restricting attention to a frequency band of special relevance); in supporting multimodal analysis of brain connectivity combining fMRI and EEG/MEG data; and in exploring changes in frequency-dependent connectivity related to drug treatments or pathological states.

Here we aim simply to provide a technical introduction to methods in the Fourier and wavelet domains that are appropriate to frequency-dependent analysis of functional connectivity in fMRI, and to illustrate them through examples.

1 Functional Connectivity Analysis in the Fourier Domain

The first steps of a functional connectivity analysis generally involve the correction of geometrical displacements (due to involuntary head movement) and the co-registration of fMRI volumes to a standard template. The examples used here come from a sample of healthy individuals that underwent a long scanning session (2048 acquisitions, TR = 1.1 s) while lying in a "resting condition" with the eyes closed. Mean time series of 90 areas corresponding to major cortical and subcortical regions of the brain were obtained from the Automated Anatomic Labelling system (AAL; Tzourio-Mazoyer et al 2002). These multivariate sets of motion-corrected time series provided the "raw" data for the frequency-dependent functional connectivity analyses described in this chapter (Fig. 1).

1.1 A General Model Underlying the Data

Some general assumptions on the statistical model underlying the data are frequently made to allow a proper analysis of the multivariate time series in the Fourier domain. S*tationarity* of time series is one of the most usual assumptions. Under stationarity each one of the P time series is expected to fluctuate around a fixed mean value μ_p ($p: 1, \ldots P$), and to have covariances between their observations dependent solely on the relative difference in their timings, i.e., the lag between timepoints, $k = t_a - t_b$.

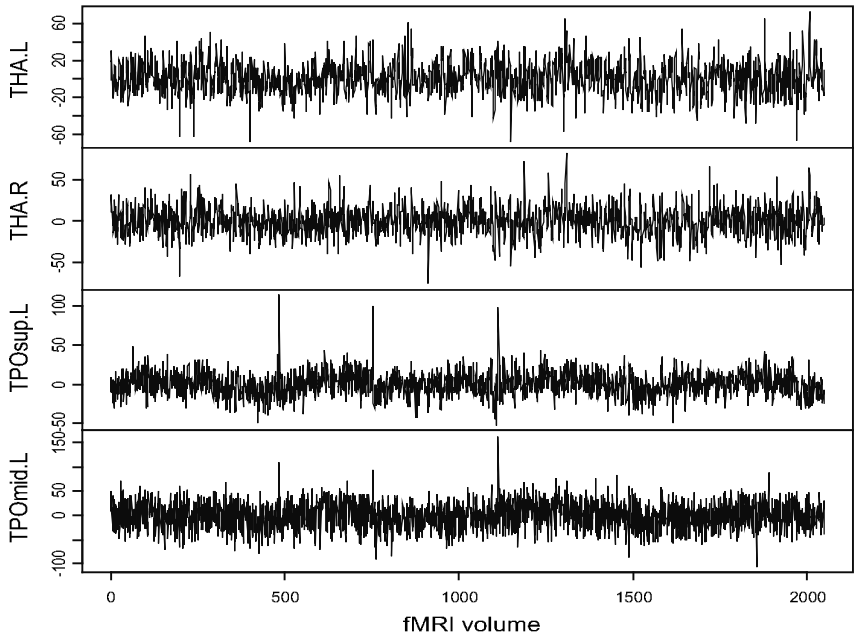


Fig. 1. Four examples of time series extracted from one subject after a long "resting state" fMRI time series. The regions shown are a subset of a parcellation that divides the brain in 90 major areas. THA.L: left thalamus, THA.R: right thalamus, TPOsup.L: left temporal pole, superior temporal gyrus, TPOmid.L: left temporal pole, middle temporal gyrus

Joint *multivariate normality* of the observations from all the time series is another assumption frequently made for convenience. The probability law of a stationary multivariate set of time series (a multivariate time series) is fully described by a vector of means $\mu = (\mu_1, \ldots, \mu_P)$ and a set of *cross-covariance functions*

$$\gamma_{p,q}(k) = Cov\left(X_p(t), X_q(t+k)\right). \tag{1}$$

For $p = q$, i.e. the observations belong to the same individual series, these functions are called auto-covariance functions. Null lag values of the auto-covariance function ($\gamma_{p,p}(0) = \sigma_p^2$) give the marginal variances of any observation of the time series; stationary covariances imply equality of variances among observations. Figure. 2 shows the estimates of the cross-covariance functions for two pairs of time series previously shown in Fig.1.

1.2 The Frequency Representation of Covariance Functions

Under the normal multivariate model, all the information on the intensity of the functional connectivity between pairs or groups of regions will be contained in the auto- and cross-covariance functions. Alternatively, the same information can be represented in the Fourier domain.

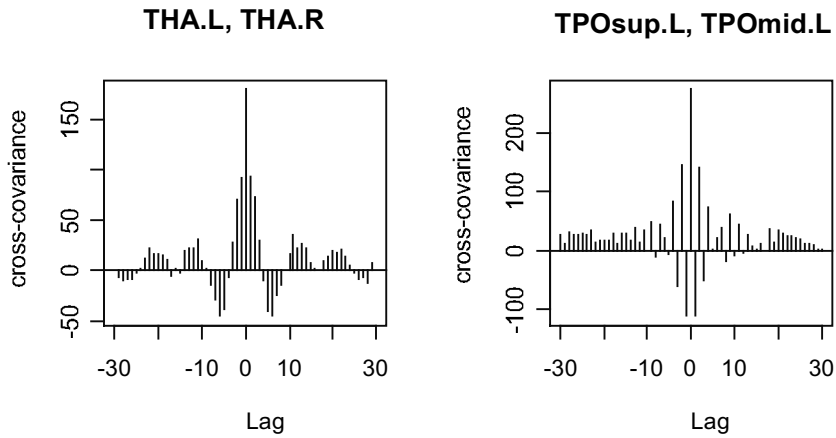


Fig. 2. Estimates of cross-covariance functions for two pairs of fMRI time series shown in Fig. 1. Standard sample cross-covariance formulas have been used (Brockwell & Davis 1991)

Encoding the covariability information in the frequency domain clarifies aspects of the data not portrayed directly by the cross-covariance. Mainly, it can be used to highlight the frequencies that subtend stronger links between two time series, and the degree of synchrony (or de-phasing) of the two signals at these specific frequencies.

The *cross-spectral density* is the counterpart of the cross-covariance function in the frequency domain. Provided that the covariances are summable ($\sum_{k=-\infty}^{\infty} |\gamma_{p,q}(k)| < \infty$) the cross-spectral density is defined by (Brillinger 1981)

$$f_{p,q}(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_{p,q}(k) e^{-ik\lambda}. \tag{2}$$

It is a continuous complex function with a periodicity of length 2π but, due to a symmetry given by the complex conjugate ($\overline{f_{p,q}(-\lambda)} = f_{p,q}(\lambda)$), it can be fully described by its values in the interval $(0,\ \pi]$. Here, using π as an upper boundary is a mathematical convention. If the time interval between two observations (the TR for fMRI) is known, the interval (0, 1/2TR] in absolute frequencies can be used alternatively.

Estimates of the cross-spectral density are usually obtained through the discrete Fourier transforms (DFT) of the original observed time series of the different brain regions, without the need to estimate their cross-covariance function. For a given Fourier frequency λ these estimates are obtained by means of a simple product of Fourier coefficients ($d_p(\lambda)$)

$$\hat{f}_{p,q}(\lambda) = d_p(\lambda)\, \overline{d_q(\lambda)} \tag{3}$$

The set of estimates at the different Fourier coefficients in the $(0, \pi]$ interval is known as the *cross-periodogram* (Brillinger 1981). The value of the product at $\lambda = 0$ is related to μ (in the time domain) and should be excluded.

However, the cross-periodogram is a rather noisy estimate of the cross-spectral density. Its variance doesn't decrease by including new observations in the time series (i.e. it is not a consistent estimator). This problem is usually solved by applying a smoothing filter over the cross-periodogram (Parzen 1961, Hannan 1970). Choosing the size of the filter will entail a trade-off between variance reduction and increased bias of the estimates. Window tapering (Tukey 1967, Brillinger 1981) over the time series, especially if these are not very long, is a recommended pre-processing step to avoid discontinuity effects on the discrete Fourier estimates. The first two rows of Fig. 3 show the moduli of the raw and filtered cross-periodograms for the two pairs of time series shown in Fig. 1.

In the same way that a standard measure of covariance does not just describe the degree of co-variability between two variables, but is also dependent on their individual variances, the cross-spectral density is also affected by the values of the individual spectral densities of each time series. Standardized measures of covariability are given by the *spectral coherence*

$$R_{p,q}(\lambda) = \frac{f_{p,q}(\lambda)}{\sqrt{f_{p,p}(\lambda)\, f_{q,q}(\lambda)}} \tag{4}$$

which has an obvious resemblance to a correlation. As with the cross-spectral density, it is a complex function but it is bounded in the complex unit circle (in the same way that correlation measures are bounded in the $[-1, 1]$ interval). Perhaps confusingly, some authors call it spectral coherency (e.g. Brillinger 1981), reserving the term spectral coherence to describe the real valued squared modulus ($|R_{p,q}(\lambda)|^2$). Sample estimates of the cross-spectral density may be obtained by plugging in the filtered cross-periodogram estimates in (4) (see some examples in Fig. 3).

When dealing with large numbers of brain regions, it may be useful to have summary measures over parts or the totality of the spectrum. This will avoid the individual analysis of all possibly generated cross-periodograms. In an information theoretical framework, the coherences can be used to define the *mutual information* between two time series (Granger & Hatanaka 1964, Brillinger 1996)

$$I_{p,q} = -\frac{1}{2\pi} \int \log \left\{ 1 - |R_{p,q}(\lambda)|^2 \right\} \, d\lambda \tag{5}$$

which, although here defined over the whole spectrum, can also be applied to specific frequency intervals of interest. The mutual information is an unbounded quantity, i.e., as coherences increase, it rises without limit. However,

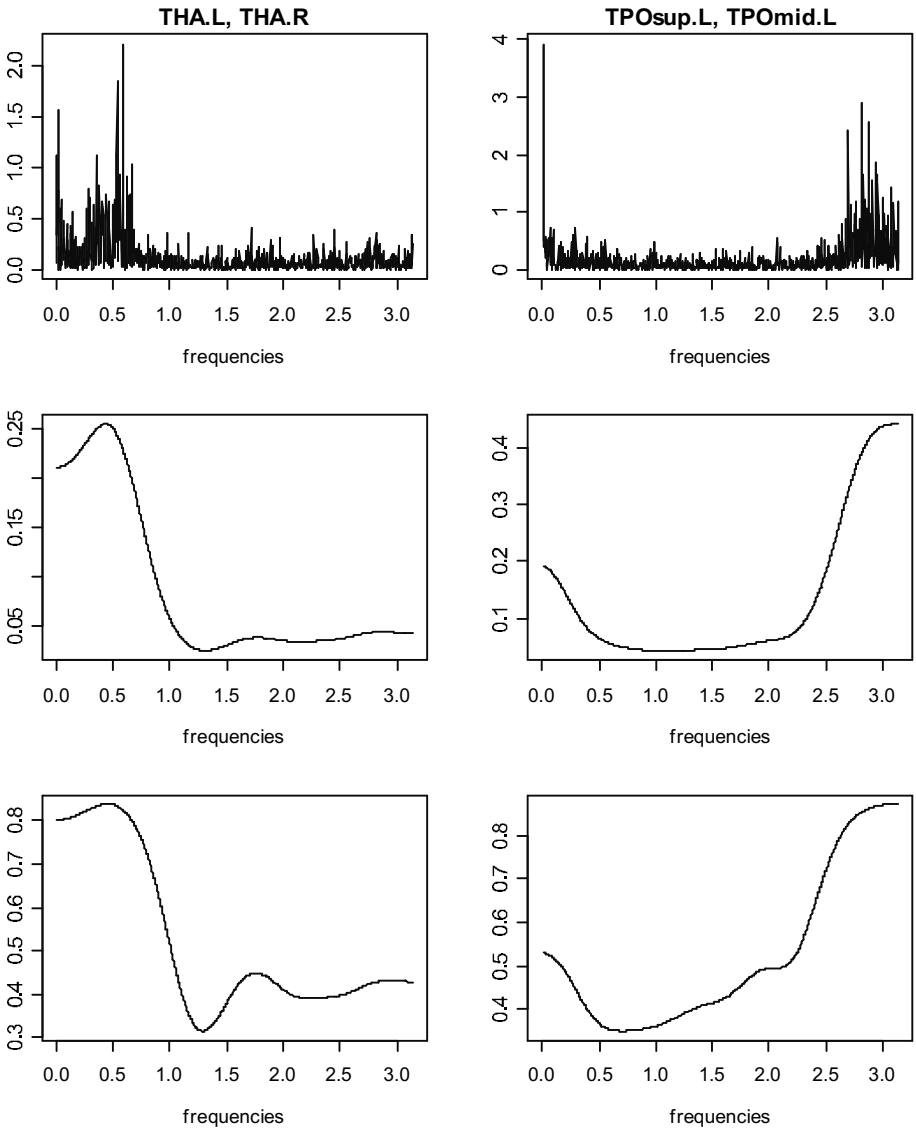


Fig. 3. Moduli of the non-filtered cross-periodograms (top row), filtered cross-periodograms (middle row), and estimated coherence functions (bottom row) of the two pairs of time series shown in Fig. 1

it can be normalised to give scores in the [0,1] interval (Granger & Lin 1994, Harvill & Ray 2000)

$$\phi_{p,q} = \sqrt{1 - \exp(-2I_{p,q})} \tag{6}$$

Estimates for both parameters can be obtained from the sample coherences by performing a finite sum over the Fourier frequencies instead of the definite integral in (5).

1.3 Conditional Measures of Association

Coherences describe the joint dynamics of two regions of the brain without taking into account their relation with the other regions included in the multivariate dataset. In that sense, they are not multivariate descriptors but just bivariate measures. Quite often one may be interested in assessing the residual covariation between two brain regions after excluding the effects on their total covariation of all other regions in the dataset, i.e., to have a conditional or partial measure of association. This may be specially appropriate for fMRI datasets where, due either to spatially spread haemodynamic patterns (Biswal et al. 1995, Lowe et al. 1998), or to movement related artifacts (Friston et al. 1996), significantly high correlations between many regions are frequently observed.

Quantities used to describe the conditional dependence between two time series in the frequency domain are developed in a similar way to those explained before for the non-conditional setting. We start by defining the conditional cross-covariance function

$$\gamma^{c}_{p,q}(k) = Cov\big(X_p(t) \big| X_{q+1}(u), \ldots, X_P(u)\ ,\ X_q(t+k) \big| X_{q+1}(u), \ldots, X_P(u)\big) \tag{7}$$

where $X_{q+1}, \ldots,\ X_P$ are the rest of the time series apart from p and q, and u is the set of all integers. Under multivariate normality, (7) will be equivalent to a partial cross-covariance function, which is the covariance function of the residuals of a least squares optimal linear filter (see Dahlhaus 2000 and Eichler 2005a)

$$\gamma^{p}_{p,q}(k) = Cov\left(\varepsilon_p(t), \varepsilon_q(t+k)\right). \tag{8}$$

From (8) we can now define the *partial cross-spectral density* ($f^{p}_{p,q}(\lambda)$) and the *partial coherence* ($R^{p}_{p,q}(\lambda)$) using formulas analogous to (2) and (4) above.

Inference on these quantities, however, is not straightforward. No estimates are directly available for the residuals of (8) under the broad normal multivariate model. This problem is overcome by an equality relating the partial coherences to the inverse of the matrix of cross-spectral densities from all pairs of regions at frequency λ (here symbolized by $g(\lambda)$) (Dahlhaus 2000)

$$R^{p}_{p,q}(\lambda) = \frac{-g_{p,q}(\lambda)}{\sqrt{g_{p,p}(\lambda) g_{q,q}(\lambda)}}. \tag{9}$$

Estimates of this parameter may be obtained by plugging in the values of the filtered cross-periodograms in (9). Two examples of partial coherence functions are given in Fig. 4.

The argument of the complex value $R^{p}_{p,q}(\lambda)$ conveys information on the degree of synchronicity or phase relations between time series p and q at frequency λ (Brillinger 1981). Partial coherence functions, however, can be transformed back to the temporal domain (as a *partial cross-correlation*

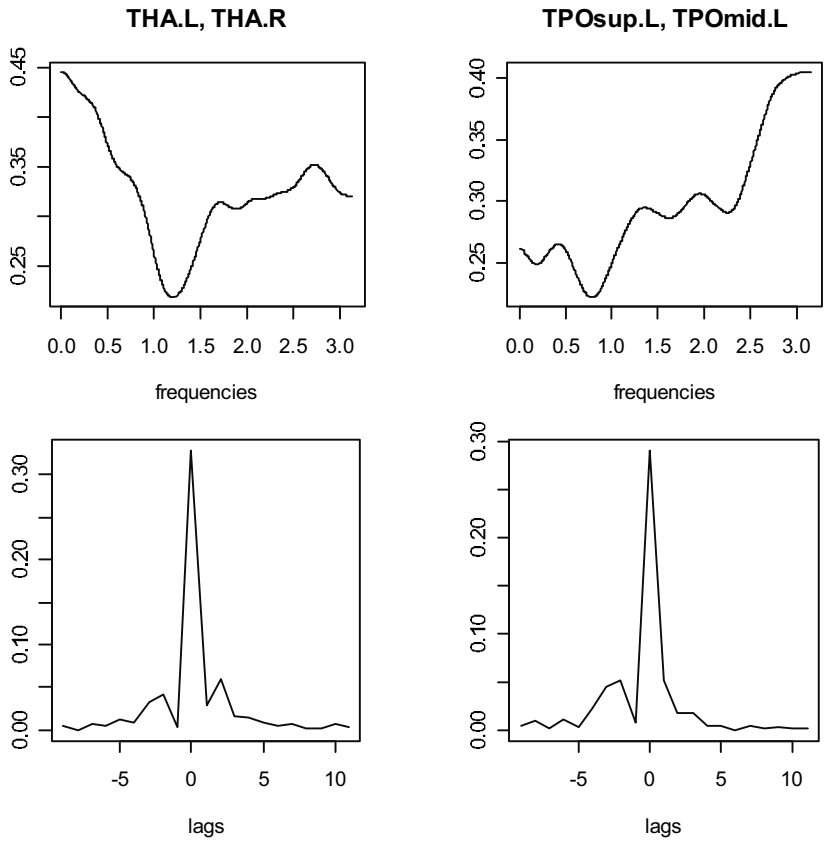


Fig. 4. Top row shows the moduli of estimated partial coherences between the two pairs of time series of Fig. 1. These are clearly different to the coherence functions shown in bottom row of Fig. 3. Bottom row shows the plots of sample partial cross-correlation functions derived from the estimated partial coherence functions. There is a clear dominance of values at zero lag (a synchronous connectivity)

function) allowing an easier visualization of time shifts in connectivity patterns. Complete explanations on how to derive the partial correlation functions from partial coherences are given in Eichler et al. (2003) and Salvador et al. (2005a). The bottom row of Fig. 4 shows the two partial correlation functions obtained from the partial coherence estimates plotted in the same figure.

Finally, when dealing with a large number of pairs of brain regions, it may be more adequate to give summary measures of conditional association over all or parts of the spectrum. For this purpose, quantities analogous to those given for the marginal coherences (5 and 6) can be simply defined for the partial coherences.

1.4 Undirected Graphs

Based on these continuous measures of frequency-dependent functional connectivity between regions it is possible to construct undirected graphs (Whittaker 1990, Cowell et al 1999), which will represent the topology of connections as a brain functional network. An *undirected graph* (G) is a mathematical object fully characterized by a set of vertices $V = \{V_1, \ldots, V_P\}$ and a set of undirected edges $E = \{(a, b) \in VxV\}$ connecting some of these vertices. In our connectivity framework, V would define the set of brain regions included in the analysis and E would contain all relevant connections between regions.

Undirected graphs are specially appropriate to portray conditional independences. Under a multivariate normal time series model, Dahlhaus (2000) defines the conditional independence between two time series by

$$R^p_{p,q}(\lambda) = 0 \quad \forall\, \lambda \in (0, \pi]. \tag{10}$$

That is, it implies having null partial coherences for all frequencies of the spectrum. Conditional independence will be depicted by an absent edge between regions p and q in the graph; conditional dependence will be represented by the presence of an edge between vertices. Besides portraying such pairwise conditional relations, undirected graphs convey information on the conditional independence of any two subsets of variables within the graph (what is known as the *global Markov property* [see Speed & Kiiveri 1986 and Cowell et al. 1999 for a more formal description]).

In real applications, the sample estimates of the partial coherences will never be zero, and some kind of decision rule will have to be applied to select the edges to include in the graph. Figure 5 shows the undirected graphs obtained by thresholding the normalized values of the partial mutual information. Values used were averages over 5 individuals scanned while resting in the scanner (Salvador et al. 2005a). Graphs are shown for low (0.0004–0.1518 Hz) and high (0.3032–0.4545 Hz) frequency ranges separately. In accordance with previous resting state studies, the amount and strength of connections is higher in the low frequencies (Cordes et al. 2001, Robouts et al. 2003), and there is an apparent symmetry between left and right hemisphere homologous regions (Lowe et al. 1998, Greicius et al. 2003). We can also show that, for most pairs of brain regions, partial coherences tend to fall off rapidly as a function of increasing anatomical distance between regions; although some pairs, such as bilaterally homologous regions of right and left cerebral hemispheres, show strong functional connections despite being separated by large distances (see Fig. 6).

1.5 Considerations on Fourier Domain Analysis of Functional Connectivity

The use of the Fourier basis to assess functional connectivity has several advantages. It allows us to find the range of frequencies subtending stronger

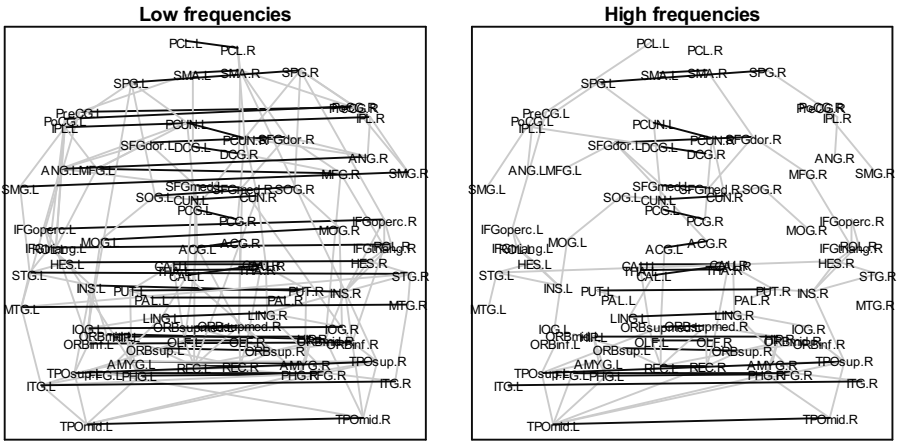


Fig. 5. Undirected graphs obtained by thresholding averaged values of the normalized partial mutual information between brain regions, for low and high frequencies separately. More connections are observed in the low frequencies. Black edges belong to inter-hemispheric symmetric pairs, and the rest of selected edges are shown in grey. The graphs are a 2D projection of the brain volume in the coronal plane; the top left quadrant of each panel representing the superior part of the left hemisphere, and so on; regional abbreviations are listed in full by Salvador et al (2005a). Reproduced with permission from Salvador et al (2005a) © Royal Society

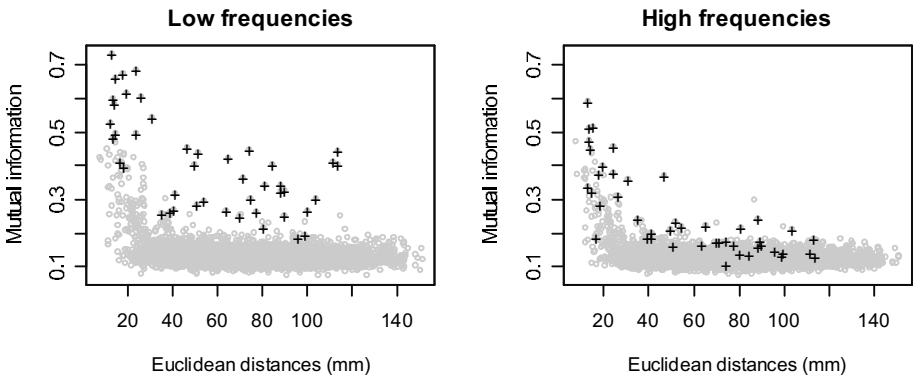


Fig. 6. Plots of normalised partial mutual information, as a function of Euclidean distances between brain regions, are shown separately for both high and low frequencies. The strength of links drops dramatically with distance for most pairs except for the predominantly inter-hemispheric symmetric couples (black crosses) at low frequencies. Reproduced with permission from Salvador et al (2005a) © Royal Society

functional links between regions and to assess the synchronicity of such connections. Furthermore, conditional measures of association may be quickly derived and portrayed visually through undirected graphs.

Some specific points should be kept in mind, though, for a proper application of the theory in functional connectivity. Although the concept of a partial correlation is rather easy to grasp, it may lead to counterintuitive results. For instance, two variables may be marginally independent but be conditionally dependent (through what is known as Berkson's paradox). In addition, the presence of latent variables (variables not included in the analysis) may have a strong effect on the connections depicted. The latter situation may be quite common in any fMRI connectivity study, where some of the neurophysiologically relevant factors may not have a spatially explicit location, or may simply not be included in the dataset. Some unavoidable degree of arbitrariness when using regions of interest will heighten this problem. The combination of marginal and conditional measures of association may help in discerning those brain connections driven by latent variables (Eichler, 2005b).

Finally, there is still a need for methodological developments. Although a significant amount of work has been done to derive the statistical properties of the parameters described here (see, for instance, Brillinger 1981, Dahlhaus et al. 1997, Dahlhaus 2000, Eichler 2004 for asymptotic distributions and for non-parametric tests, and Fried & Didelez 2003 for model selection procedures), further improvements in different aspects, such as the problem of multiple comparisons, or the development of hierarchical models to assess differences between groups may be required.

2 Functional Connectivity Analysis in the Wavelet Domain

As with the Fourier techniques, where a given signal is described as a weighted sum of sine and cosine functions, wavelet techniques use another set of basis functions (the wavelets) with a similar objective. However, while the Fourier basis gives a precise frequency localization, wavelets have the advantage of decomposing the signal in both time and scale (each scale being equivalent to a frequency band; see the example of Table 1). Such dual decomposition is achieved by using filter functions with local scope (Percival & Walden 2000); see Bullmore et al (2004) for a review of wavelets in relation to fMRI data analysis.

Although the discrete wavelet transform (DWT) has been the most frequently applied method to decompose time series on a scale by scale basis, in this chapter we will use the Maximal Overlap DWT (MODWT), as it is better suited to derive wavelet based connectivity measures and their estimates (Whitcher et al. 2000, Gençay et al. 2002).

Wavelet transforms are applied through linear filters on the original time series. Different filters are used to obtain the wavelet coefficients at different

Table 1. Values of the main 'small world' quantities for each scale obtained when the maximum permissible thresholds (R) were applied to wavelet correlation matrices representing human brain functional connectivity. All quantities have been defined in the main text. f is the frequency interval spanned by each wavelet scale

Scale	f (Hz)	R	L_{net}	C_{net}	λ	γ	σ
1	0.23–0.45	0.13	2.9	0.534	1.28	1.81	1.42
2	0.11–0.23	0.2	2.6	0.566	1.12	2.2	1.97
3	0.06–0.11	0.39	2.69	0.555	1.16	2.26	1.95
4	0.03–0.06	0.44	2.49	0.525	1.09	2.37	2.18
5	0.01–0.03	0.35	2.4	0.554	1.05	2.35	2.25
6	0.007–0.01	0.17	2.65	0.515	1.14	2.15	1.87

scales. Each scale (τ_j) is related to a time resolution (and to its associated frequencies) by $\tau_j = 2^{j-1}$ but, for simplicity, scales are frequently identified through their scale index j.

For a time series of length N, and a chosen primary wavelet filter of length L, a filter h_j for scale τ_j with $L_j = (2^j - 1)(L - 1) + 1$ coefficients will be defined. Next, MODWT filters are derived by rescaling h_j, and by circularly shifting the rescaled filters by integer units to the right (see Percival & Walden 2000 for an exhaustive description of the MODWT filters and their implementation). Fig. 7 shows the N-dimensional wavelet coefficients for three different scales, obtained by applying the MODWT filters on a fMRI time series of the left thalamus.

The main advantage of wavelet filters over Fourier filters, in a functional connectivity framework, is their ability to deal with a broader range of time series. Stationary time series with non-summable covariances $\sum_{k=-\infty}^{\infty} |\gamma_{i,j}(k)| = \infty$, known as long range dependent processes (Abry & Veitch 1998), can not be properly analyzed with Fourier methods, since the spectral density is unbounded for some of the frequencies. This is a pertinent issue for fMRI data analysis because these time series are often characterized by such long-memory properties (Maxim et al. 2005). Furthermore, the use of wavelets can be extended to non-stationary models. Specifically, they are also appropriate for the general family of d-backward difference stationary models (Hosking 1981) which do have a covariance function, although it cannot be easily estimated due to the non-stationarity of the mean. An appropriately chosen initial wavelet filter (with length $L > 2d$) will return zero mean stationary coefficients, and will allow proper estimation of connectivity measures between non-stationary time series based on their wavelet coefficients.

2.1 Measures of Association Based on the Wavelet Coefficients

Several measures of association between two time series have been derived from the wavelet coefficients (see Whitcher et al. 2000 and Gençay et al. 2002).

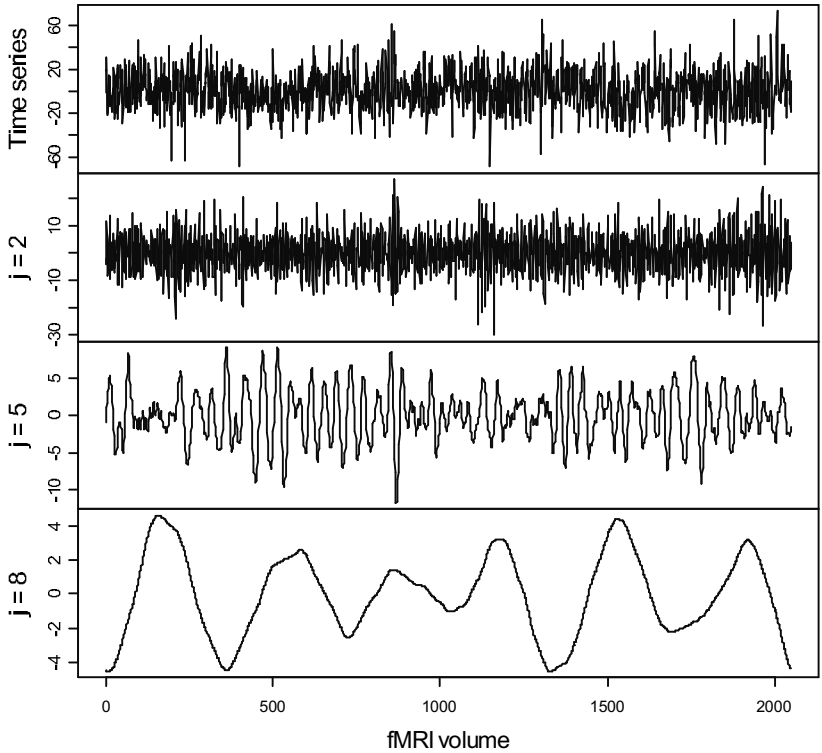


Fig. 7. Time series for the left thalamus together with the series of derived MODWT wavelet coefficients for three different scales ($\tau_2 = 2,\ \tau_5 = 16,\ \tau_8 = 128$). As the scale index j increases the time span increases and the wavelet coefficients describe lower frequency patterns of the original series

Given two d-backward differentiable stochastic processes, $X_1(t)$ and $X_2(t)$, with stationary MODWT coefficients $W_{1,j}(t)$ $W_{2,j}(t)$ at scale $\tau_j = 2^{j-1}$, the wavelet covariance is defined by

$$\gamma_{W_1,W_2}(\tau_j) = \frac{1}{2\tau_j} Cov(W_{1,j}(t), W_{2,j}(t)) \tag{11}$$

which, due to the stationarity of the coefficients, will be time invariant. An estimate of the wavelet covariance from the MODWT coefficients of the N-length observed time series is given by

$$\hat{\gamma}_{W_1,W_2}(\tau_j) = \frac{1}{N - L_j + 1} \sum_{t=L_j-1}^{N-1} w_{1,j}(t)\, w_{2,j}(t). \tag{12}$$

The top row of Fig. 8 plots the estimated wavelet covariances, at different scales, for the two pairs of time series in Fig. 1.

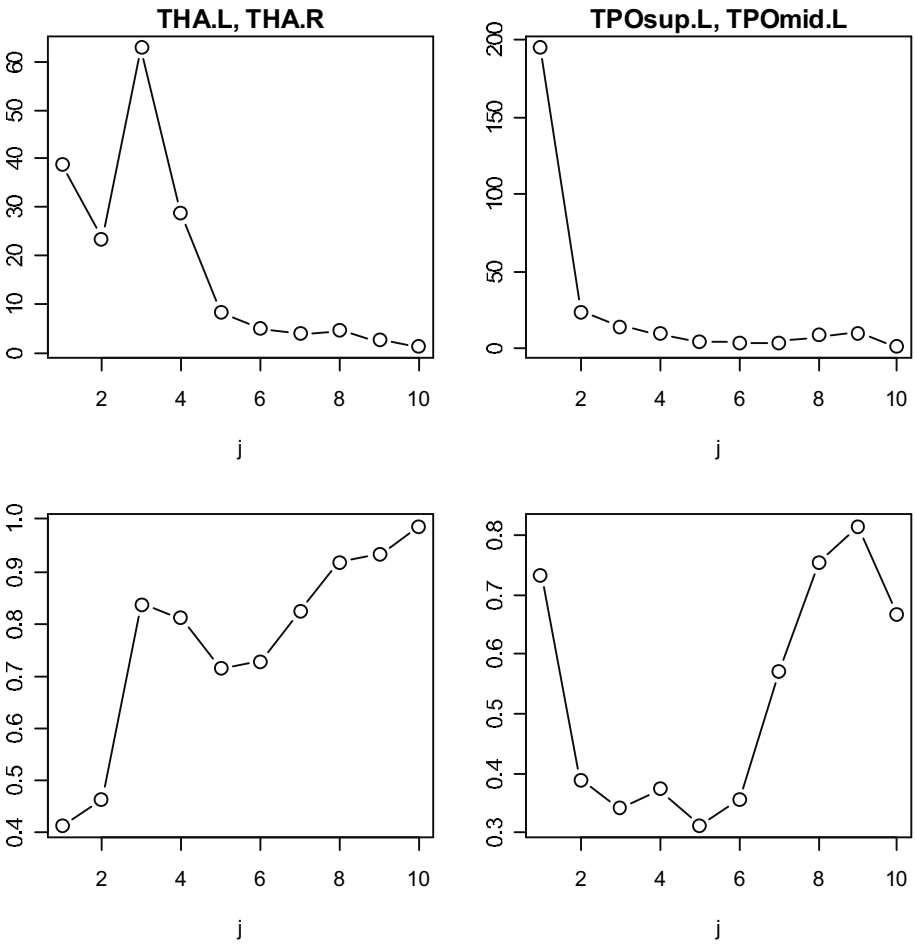


Fig. 8. Estimated wavelet covariances (top row) and wavelet correlations (bottom row) as a function of scale index j, for the two pairs of time series shown in Fig. 1. Low frequencies, represented by larger scales, account for strongest correlations between right and left pairs of thalamic and temporal polar regions

As happens with any standard covariance measure, the wavelet covariance will be affected by the intrinsic variability of both time series. This can be overcome using a standardized measure of covariability, which is given by the wavelet correlation

$$\rho_{W_1,W_2}(\tau_j) = \frac{\gamma_{W_1,W_2}(\tau_j)}{\sqrt{\gamma_{W_1,W_1}(\tau_j)\,\gamma_{W_2,W_2}(\tau_j)}} \tag{13}$$

A simple estimate of this quantity is obtained by plugging-in values of (12) (see examples in Fig. 8). Good asymptotic properties for both wavelet covariance and correlation estimates are shown in Whitcher et al. (2000) under some

general conditions. Finally, by considering lagged relations between wavelet coefficients, both $\gamma_{W_1,W_2}(\tau_j)$ and $\rho_{W_1,W_2}(\tau_j)$ can be extended to wavelet cross-covariance and cross-correlation functions, although these may be difficult to interpret and are not widely used yet.

2.2 A Small World Analysis of Wavelet Based Connectivity Graphs

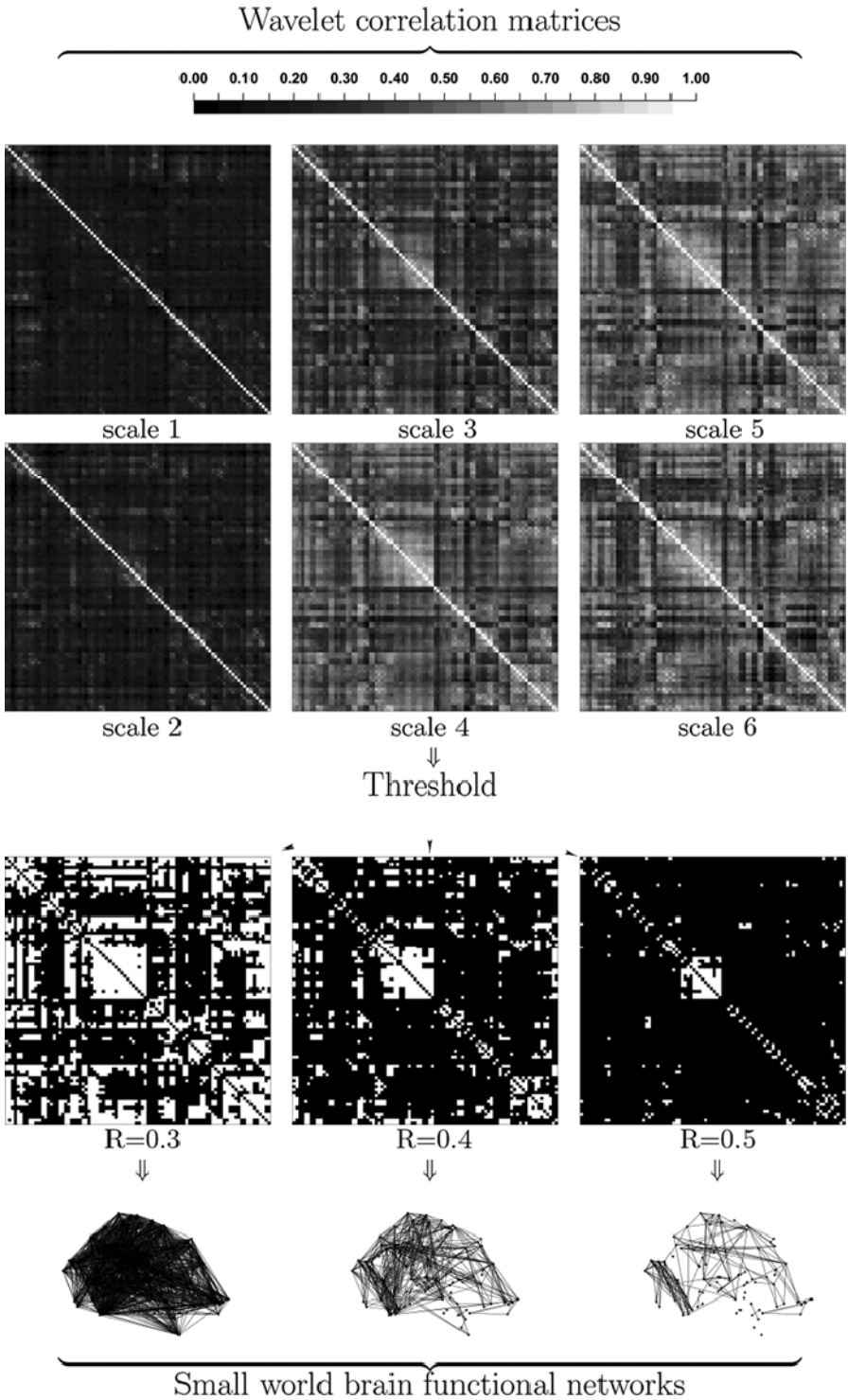
In a similar way to the frequency specific graphs derived in Sect. 1.4, the wavelet correlation matrices obtained at each scale may be used to generate an undirected graph portraying the scale specific topology of brain functional networks. Figure 9 shows the steps involved in the creation of these wavelet based connectivity graphs. After deriving the estimates of the wavelet correlation matrices from the MODWT coefficients, the correlations are thresholded to generate binary matrices coding the edges between nodes in the graph. Finally, a spatially explicit representation of these adjacency matrices leads to visualization of the graphs as brain networks in anatomical space.

Once an undirected graph of whole brain functional connectivity subtended by a specific frequency interval has been constructed in this way, we can explore its topological properties using tools drawn from small-world analysis of other complex networks (Watts & Strogatz 1998; Strogatz 2001). Small world networks are associated with a high density of local connections between nodes and a short geodesic distance between any two nodes in the network. This makes them attractive as models of brain functional architecture, which is theoretically anticipated to support both segregated/modular and distributed/global modes of information processing (Sporns and Zwi, 2004). The basic metrics required for a small world analysis are:

a) k_i, the number of undirected edges reaching vertex (brain region) V_i.
b) The clustering coefficient C_i ($0 < \mathrm{C_i} < 1$), a ratio which defines the proportion of possible connections that actually exist between the nearest neighbors of a vertex (Watts & Strogatz 1998); high values of $\mathrm{C_i}$ imply that most of the nearest neighbors of that vertex are also nearest neighbors of each other, or that the vertex is located in a cliquish local neighborhood.
c) The minimum path length, defined as the number of edges comprising the shortest path between a given pair of vertices; the mean minimum path length L_i is the average of the n-1 minimum path lengths between Vi and all other vertices in the network.

Values of these parameters can be used to identify the "hubs"; highly connected nodes with largest values of k_i or smallest values of L_i. Furthermore, all these quantities can be averaged over vertices to estimate the network means $\mathrm{k_{net}}$, $\mathrm{C_{net}}$ and the characteristic path length $\mathrm{L_{net}}$.

To diagnose small-world properties, the characteristic path length and clustering coefficient are compared to the same parameters estimated in random networks configured with the same number of vertices, mean degree



k_{ran} and degree distribution as the network of interest, under the constraint that $k_{\text{ran}} > \log(\#V)$ ($\#V$ being the number of vertices). Typically in a small-world network we expect the ratio $\gamma = C_{\text{net}}/C_{\text{ran}} > 1$ and the ratio $\lambda = L_{\text{net}}/L_{\text{ran}} \sim 1$. A scalar summary of "small-worldness" is therefore the ratio $\sigma = \gamma/\lambda$ which is typically greater than 1 (Watts & Strogatz, 1998; Montoya & Solé 2002; Humphries et al. 2005).

As shown in Fig. 10, the value of these parameters may be significantly affected by the values of the threshold used to obtain the graphs. Specifically, for the fMRI datasets analyzed, the most "small world-like" graphs (highest values of σ) were obtained with the highest possible thresholds applied to scales 3, 4 and 5, just before reaching the limit set by $k > \log(\#V)$; see also Achard et al. (2006). Table 1 gives the values of the parameters at the 6 different scales when the highest allowable thresholds were applied.

Finally, the analysis of the distribution of the k_i values for all vertices may give some additional clues on the properties of the brain connectivity graphs. Figure. 11 shows the histogram of k_i values for scale 4, together with its degree of fit to different theoretical models. A truncated power law model for the cumulative degree distribution seems to be the most appropriate for this data. Similar forms for the degree distribution have previously been reported for diverse biological and other networks and typically indicate the existence of important physical constraints (such as aging of nodes or connection costs) on emergence of very highly connected hubs (Strogatz, 2001, Amaral 2000).

Conclusions

We have described and illustrated Fourier and wavelet methods of frequency-dependent functional connectivity analysis applied to human fMRI data. We have shown how analysis of multivariate time series in both domains can support construction of undirected graphs representing sparse but complex whole brain functional networks; and in the wavelet domain we have shown that in the framework of the small-world theory, these networks are not scale free.

Fig. 9. Plot showing the main steps needed to create wavelet based connectivity graphs. Functional MRI time series from the different regions are filtered to MODWT coefficients. For each scale, the coefficients are used to estimate the wavelet correlation matrix (matrices shown here are averages from five different individuals scanned under resting conditions). Next, a threshold parameter (R) is used to create the binary matrices that will code for the edges in the spatially explicit undirected graphs. As shown for scale 4, different values of the threshold parameter will lead to graphs of different sparsities. The typically small-world combination of high clustering (e.g. in occipital cortex) and short path length mediated by some long-range connections (e.g. between frontal and parietal association cortex) is evident by inspection of the more conservatively thresholded graphs. Reproduced with permission from Achard et al (2006) © Society for Neuroscience

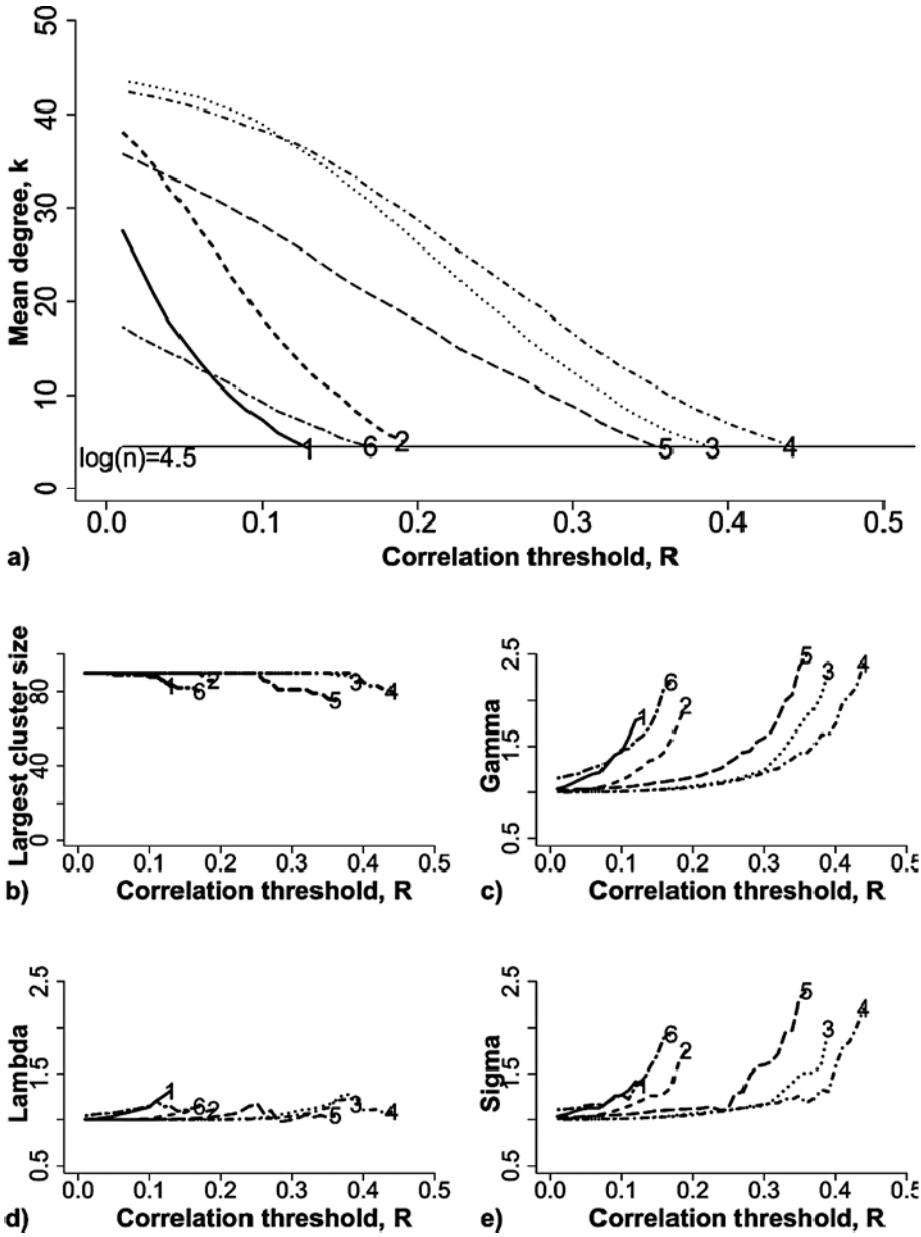


Fig. 10. Plots showing the values of some of the small world quantities described in the text as a function of the thresholds applied to the wavelet correlation matrices of Fig. 9. Each line refers to a different scale. Both mean degree and maximum cluster size are decreased by increasing thresholds. While $\lambda = L_{net}/L_{ran}$ values remained stable around one, $\gamma = C_{net}/C_{ran}$ values were increased with R, leading to an overall increase in the small-worldness scalar $\sigma = \gamma/\lambda$ with values greater than 2 for low frequency scales at high thresholds. Reproduced with permission from Achard et al (2006) © Society for Neuroscience

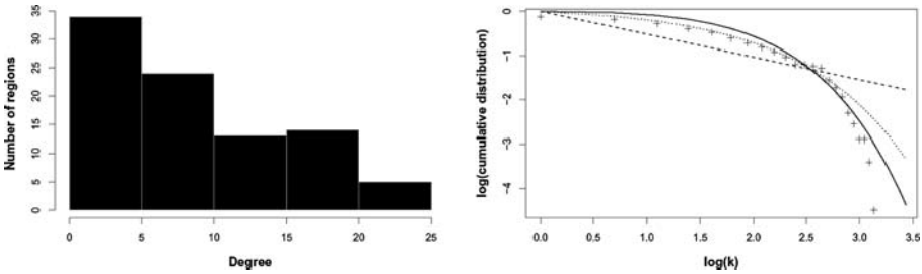


Fig. 11. Degree distribution of a small-world brain functional network. Left plot: histogram of regional degree k_i distribution. Right plot: plot of the log of the cumulative probability of degree, $\log(\mathrm{P}(k_i))$, versus log of degree, $\log(k_i)$. The plus sign indicates observed data, the solid line is the best-fitting exponentially truncated power law, the dotted line is an exponential, and the dashed line is a power law. Reproduced with permission from Achard et al (2006) © Society for Neuroscience

Acknowledgments

This work has been supported by a grant of the Spanish Health Ministry (CP04/00322) to Raymond Salvador. Software development was supported by a Human Brain Project grant from the National Institute of Mental Health and the National Institute of Biomedical Imaging and Bioengineering. The Wolfson Brain Imaging Centre was supported by an MRC Co-operative Group grant.

References

Abry P, Veitch, D (1998), Wavelet analysis of long range dependent traffic, IEEE Transactions on Information Theory 44, 2–15.

Achard S, Salvador R, Whitcher B, Suckling J, Bullmore E (2006) A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. J Neurosci 26, 63–72.

Amaral LAN, Scala A, Barthélémy M, Stanley HE (2000) Classes of small-world networks. Proc Natl Acad Sci USA 97, 11149–11152.

Biswal B, Yetkin FZ, Haughton VM, Hyde JS (1995) Functional connectivity in the motor cortex of resting human brain using echoplanar MRI. Magn Reson Med 34, 537–541.

Brillinger DR (1981) Time Series. Data analysis and theory. 2rd ed., Holden Day, San Francisco.

Brillinger, DR (1996) Remarks concerning graphical models for the time series and point processes. Rev. Econ. 16, 1–23.

Brockwell PJ, Davis RA (1991) Time Series. Data analysis and theory. 2rd ed., Springer, New York.

Bullmore ET, Fadili J, Maxim V, Sendur L, Whitcher B, Suckling J, Brammer M, Breakspear M (2004) Wavelets and functional magnetic resonance imaging of the human brain. NeuroImage 23, 234–249.

Bullmore ET, Rabe-Hesketh S, Morris RG, Williams SC, Gregory L, Gray JA, Brammer MJ (1996) Functional magnetic resonance image analysis of a large-scale neurocognitive network. Neuroimage 4, 16–33.

Cordes D, Haughton VM, Arfanakis K, Carew JD, Turski PA, Moritz CH, Quigley MA, Meyerand ME (2001) Frequencies contributing to functional connectivity in the cerebral cortex in 'resting state' data. Am J Neuroradiol 22, 1326–1333.

Cowell RG, Dawid AP, Lauritzen SL, Spiegelhalter DJ (1999) Probabilistic networks and expert systems. Springer, New York.

Dahlhaus R (2000) Graphical interaction models for multivariate time series. Metrika 51, 157–172.

Dahlhaus R, Eichler M, Sandkühler J (1997) Identification of synaptic connections in neural ensembles by graphical models. Journal of Neuroscience Methods 77, 93–107.

Eichler M (2004) Testing nonparametric hypothesis in non-Gaussian vector stationary time series. Techical Report, Universität Heildelberg.

Eichler M (2005a) Fitting graphical interaction models to multivariate time series. Techical Report, Universität Heildelberg.

Eichler M (2005b) A graphical approach for evaluating effective connectivity in neural systems. Phil. Trans. R. Soc. B 360, 953–967.

Eichler M, Dahlhaus R, Sandkühler J (2003) Partial correlation analysis for the identification of synaptic connections. Biol. Cybern. 89, 289–302.

Espinosa IE, Gerstein GL (1988) Cortical auditory neuron interactions during presentation of 3-tone sequences: effective connectivity. Brain Res 450,39–50.

Fried R, Didelez V (2003) Decomposability and selection of graphical models for time series. Biometrika 90, 251–267.

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6, 218–29.

Friston KJ, Williams SR, Howard R, Frackowiak RSJ, Turner R (1996) Movement-Related Effects in fMRI Time-Series. Magn Reson Med 35, 346–355.

Gençay R, Selçuk F, Whitcher B (2002) An introduction to wavelets and other filtering methods in finance and economics. Academic press, San Diego.

Granger, CWG, Hatanaka M (1964) Spectral analysis of economical time series. Princeton Press, Princeton.

Granger C, Lin JL (1994) Using the mutual information coefficient to identify lags in nonlinear models. J. Time Ser. Anal. 15, 371–384.

Greicius MD, Krasnow B, Reiss AL, Menon V (2003) Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc Natl Acad Sci USA 100, 253–258.

Hannan EJ (1970) Multiple time series. Wiley, New York.

Harvill J, Ray B (2000) Lag identification for vector nonlinear time series. Commun. Stat. Theory Methods 29, 1677–1702.

Horwitz B (2003) The elusive concept of brain connectivity. Neuroimage 19,466–470.

Hosking JRM (1981) Fractional differencing. Biometrika 68, 165–176.

Humphries MD, Gurney K, Prescott TJ (2005) The brainstem reticular formation is a small-world, not scale-free, network. Proceedings of the Royal Society B. In press. DOI:10.1098/rspb.2005.3354

Koenig T, Studer D, Hubl D, Melie L, Strik WK (2005) Brain connectivity at different time-scales measured with EEG. Phil. Trans. R. Soc. B. 360,1015–1023.

Leopold DA, Logothetis NK (2003) Spatial patterns of spontaneous local field activity in the monkey visual cortex. Rev Neurosci 14,195–205.

Lowe MJ, Dzemidzic M, Lurito JT, Mathews VP, Phillips MD (2000) Correlations in low-frequency BOLD fluctuations reflect cortico-cortical connections. Neuroimage 12, 582–587.

Lowe MJ, Mock BJ, Sorenson JA (1998) Functional connectivity in single and multislice echoplanar imaging using resting state fluctuations. Neuroimage 7, 119–132.

Maxim V, Sendur L, Fadili J, Suckling J, Gould R, Howard R, Bullmore E (2005) Fractional Gaussian noise, functional MRI and Alzheimer's disease. Neuroimage 25,141–158.

Mayhew JE, Askew S, Zheng Y, Porrill J, Westby GW, Redgrave P, Rector DM, Harper RM (1996) Cerebral vasomotion: a 0.1-Hz oscillation in reflected light imaging of neural activity. Neuroimage 4,183–93.

McIntosh AR (1999) Mapping cognition to the brain through neural interactions. Memory 7,523–548.

Montoya JM, Solé RV (2002) Small world patterns in food webs. J. Theoret. Biol. 214, 405–412.

Parzen E (1961) Mathematical considerations in the estimation of spectra. Technometrics 3, 167–190.

Percival DB, Walden AT (2000) Wavelet methods for time series analysis. Cambridge University Press, Cambridge.

Ramnani N, Behrens TEJ, Penny W, Matthews PM (2004) New approaches for exploring Anatomical and functional connectivity in the human brain. Biol Psychiatry 56, 613–619.

Rombouts SARB, Stam CJ, Kuijer JPA, Scheltens P, Barkhof F (2003) Identifying confounds to increase specificity during a 'no task condition': evidence for hippocampal connectivity using fMRI. Neuroimage 20, 1236–1245.

Salvador R, Suckling J, Coleman MR, Pickard JD, Menon D, Bullmore E (2005b) Neurophysiological architecture of functional magnetic resonance images of human brain. Cereb Cortex. 15,1332–1342.

Salvador R, Suckling J, Schwarzbauer C, Bullmore E (2005a) Undirected graphs of frequency-dependent functional connectivity in whole brain networks. Phil. Trans. R. Soc. B 360, 937–946.

Samar VJ, Swartz KP, Raghuveer MR (1995) Multiresolution analysis of event-related potentials by wavelet decomposition. Brain Cogn 27,398–438.

Speed TP,. Kiiveri HT (1986) Gaussian Markov distributions over finite graphs. Ann. Stat. 14, 138–150.

Sporns O, Zwi JD (2004) The small world of the cerebral cortex. Neuroinformatics 2,145–162.

Strogatz SH (2001) Exploring complex networks. Nature 410, 268–276.

Sun FT, Miller LM, D'Esposito M (2004) Measuring interregional functional connectivity using coherence and partial coherence analyses of fMRI data. Neuroimage 21,647–658.

Tukey (1967) An introduction to the calculations of numerical spectrum analysis. In: Harris B (ed.) Advanced seminar on spectral analysis of time series, Wiley, New York, pp 25–46.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273–289.

Watts DJ, Strogatz SH (1998) Collective dynamics of "small-world" networks. Nature 393, 440–442.

Whitcher B, Guttorp P, Percival DB (2000). Wavelet analysis of covariance with application to atmospheric time series. Journal of Geophysical Research, 105, 941–962.

Whittaker J (1990) Graphical models in applied multivariate statistics. Wiley, Chichester.

Yamashita O, Sadato N, Okada T, Ozaki T (2005) Evaluating frequency-wise directed connectivity of BOLD signals applying relative power contribution with the linear multivariate time-series models. Neuroimage 25,478–90.

Part IV

Behavioral, Developmental and Clinical Applications

The Role of Neural Context in Large-Scale Neurocognitive Network Operations

Steven L Bressler[1] and Anthony R McIntosh[2]

[1] Center for Complex Systems & Brain Sciences, Florida Atlantic University, Boca Raton FL 33431 USA
[2] Rotman Research Institute, Baycrest Centre, Toronto ON Canada M6A 2E1

Brain connectivity has been a central factor in the development of theories about the mind-brain link. In its simplest form, brain connectivity analysis has revealed serial processing systems, wherein specific neural elements (neurons, neuronal assemblies, neuronal populations) cooperate to express a circumscribed function that is realized as information passes through the system in a feedforward manner. Consideration of parallel architectures provides a more complex view of system processing by revealing that each brain region may impact many other regions through direct and indirect routes, including areas from which it receives its input. Regardless of the specific architecture, the notion that cognition results from the operations of large-scale neural networks has been present in various forms throughout the history of neuroscience (Finger, 1994; Bressler 1995, 2002). For a large part of that history, it was difficult to verify this notion because most available methods only allowed investigation of small parts of the nervous system in isolation. Ideally, simultaneous measures from many parts of the brain must be analyzed to understand the operations of large-scale networks that underlie cognition. In the past few decades, advances in functional neuroimaging, including Positron Emission Tomography (PET) functional Magnetic Resonance Imaging (fMRI), and EEG/MEG-based source localization, have allowed simultaneous distributed measures of brain function to be related to cognition.

This chapter examines the role of a critical aspect of brain function, which we call *neural context*, in the selective functioning of interacting neural systems in cognition. We define neural context as the local processing environment of a given neural element that is created by modulatory influences from other neural elements. Neural context allows the response properties of one element in a network to be profoundly affected by the status of other neural elements in that network. As a result of neural context, the relevance of a given neural element for cognitive function typically depends on the status of other interacting elements (McIntosh 1999; Bressler 2003a). By this definition, the processing performed by a given brain area may be modulated by a potentially

large number of other areas with which it is connected. Since brain areas are most often bidirectionally connected, the neural context of each connected area emerges spontaneously from its interactions. Furthermore, to the extent that basic sensory and cognitive operations share similar brain constituents, they experience similar neural contextual influences.

Neural context refers only to the context that arises within the brain as a result of interactions between neural elements. In this chapter, we also distinguish a related form of context, which we refer to as *situational context*. Unlike neural context, situational context represents a host of interrelated environmental factors, including aspects of the sensory scenes and response demands of both the external and internal milieus. A red light presented to a person in isolation usually means nothing, but a red light presented to that person while driving elicits a situationally specific response. Situational context is most often what researchers have in mind when they examine "contextual effects" on the brain (Hepp-Reymond et al., 1999; Chun 2000; Herzog et al., 2002; Bar 2004; Beck & Kastner 2005).

In most normal circumstances, neural context is shaped by situational context. The environments in which animals and humans must survive have a high degree of structural complexity, an important consequence of which is a fundamental uncertainty in the organism's perceptuo-motor interactions with those environments. Complete information is never available to allow total certainty about the state of the environment and the optimal course of action in it. The limited information that the organism has about its environmental situation usually renders ambiguous its perceptual interpretation of environmental entities and the appropriate actions to be directed toward them. The ability to utilize and manipulate information about the organism's situational context, can dramatically reduce uncertainty, thereby enhancing the organism's interactions with the environment and lending survival advantage to its species.

Since the complexity of the environment's structure spans multiple structural and temporal scales, situational context must affect all types of cognitive function, including sensation, perception, emotion, memory, planning, decision making, and action generation. It is reasonable to infer therefore that neural context should also be of primary importance in the implementation of those functions by the brain. In other words, just as situational context can have effects at multiple scales and across multiple behaviors, so too is neural context expected across all spatial and temporal scales in the brain and across all behaviors.

1 Anatomical and Physiological Foundations of Neural Context

A fundamental factor that supports contextual processing in the brain is its large-scale connectivity structure. The anatomical connectivity of the cerebral cortex, in particular, appears to have evolved to support contextual

processing. The cortex consists of a large number of areas profusely interconnected in a complex topological structure, which places strong constraints on its functional dynamics (Sporns et al., 2000; Bressler & Tognoli 2006; Sporns & Tononi 2007). In sensory systems, local cortical networks are interconnected by feedforward, feedback, and lateral connections (Felleman & Van Essen 1991), all of which may serve to provide neural context for the processing that occurs in a given local network. Integration along convergent feedforward pathways from peripheral receptor sheets may be sufficient for some forms of sensory contextual processing, whereas other forms may require lateral and feedback connections. Contextual processing in cortical motor and association areas also critically depends on the complex patterning of interconnected local networks (Brovelli et al., 2004).

Although the local networks in different cortical areas show cytoarchitectonic variation, the cellular components and internal connectivity of cortical circuits are generally similar throughout the cortex. What distinguishes the specialized function of any local cortical network is its topological uniqueness, i.e. its particular pattern of interconnectivity with other networks. The unique set of local networks with which a given local cortical network is directly connected has been called its "connection set" (Bressler 2002, 2003a) or "connectional fingerprint" (Passingham et al., 2002). Providing direct synaptic input to the circuit elements of the local network, the networks of the connection set have privileged status in creating context for the processing in that local network. The connection set of a local network thus determines the contextual guidance that the network receives during processing, and consequently modulates the trajectory of the local processing dynamics.

A second factor in cortical contextual processing is spatial pattern formation in local cortical networks (Beggs et al., 2007). The generation and transmission of spatially patterned activity by local networks is central to interarea communication in the cortex (Freeman 2003; Andras 2005), and provides a realistic framework for contextual modulation. From this viewpoint, the processing dynamics in a local cortical network is manifested by the formation of spatially patterned activity under modulation by spatial activity patterns transmitted from the members of its connection set (Bressler 2004). Neural context is thereby implemented through the interaction of transmitted spatial activity patterns from the connection set with patterns generated by the local circuitry. Transmission uses parallel, convergent, and divergent pathways between transmitting and receiving networks (Fuster 2003).

A third aspect of cortical function necessary for contextual processing is reentrant interaction (Tononi et al., 1992). Since the anatomical interconnection of cortical areas is largely bidirectional (Felleman & Van Essen 1991), a local network in one area receives feedback from the same local networks to which it transmits in other areas, i.e. transmitted influences are reciprocated by reentrant influences from receiving networks. Reentrant interactions that a local network undergoes with its connection set modulate its spatial pattern

processing, resulting in the alteration of its generated patterns. Reentrant interactions thereby enrich the processing capability of the local circuitry.

Natural environments are rich with information about situational context. The capacity to utilize that information enhances the behavioral adaptation of animals to their surroundings. The exploitation of situational context information affords an adaptive advantage, which exerts a strong selection pressure for the evolution of neural architectures and mechanisms conducive to the efficient processing of that information. Primates appear to have evolved special behavioral prowess due to their highly developed contextual processing abilities. The behavioral capability for taking advantage of situational context information depends on the brain's faculty for processing that information. It is quite likely that the neural processing of information about situational context depends on the deployment of neural context, by which the local processing in an area of the brain is modulated by its interactions with other brain areas. To understand the use of situational context information in behavior, we consider here the implementation of neural context in the primate cerebral cortex, treating in particular some basic anatomical and physiological features that allow and promote cortical contextual processing.

It is generally agreed that the functions of local networks in circumscribed cortical areas depend on interactions among the neuronal elements of their circuitry (DeFelipe et al., 2002; Douglas & Martin 2004; Bressler & Tognoli 2006). There is also growing appreciation of the fact that cortical circuit function is expressed in real time by the dynamics of these interactions (Lopes da Silva et al., 1994; Friston & Price 2001; Siegel & Read 2001; Breakspear & Jirsa 2007). Neural context for the functional dynamics of a local cortical circuit is provided by modulations impinging on the circuit elements from other connected circuits. It may be said that the circuit dynamics undergo *contextual guidance* by such modulatory influences. In simulation studies, it has been shown that the computational capabilities of a local processor can be greatly enhanced by contextual guidance from the outputs of other processors. (Becker & Hinton 1992; Kay & Phillips 1997; Phillips et al., 1998). Contextual guidance may similarly be at work in normal cortical operations (Phillips & Singer 1997).

In our view, the neural context for processing in a local network emerges spontaneously through the reentrant interactions that the network undergoes with the members of its connection set. The modulation of spatial pattern formation that occurs in the local network as a result of those interactions constitutes contextual guidance of the local processing dynamics, constraining it to trajectories that are contextually consistent with activity in the connection set. Since reentrant interactions involve interconnected local networks distributed throughout the cerebral cortex, it is to be expected that spatial pattern formation will proceed concurrently in each local network under contextual modulation from all the networks to which it is connected. Thus, due to the massive large-scale interconnectivity of cortex, contextual effects are expected to be ubiquitous, and contextual guidance to occur concurrently in

numerous widespread local networks, each both transmitting and receiving contextual effects.

The interplay of multiple recurrent interactions across the cortex has been postulated (Bressler 2004) to lead to the emergence of a global cortical context that reflects the current situational context. This process is thought to involve the convergence of local networks to mutually consistent activity patterns that realize informational coherence and incoherence relations (Thagard & Verbeurgt 1998). Thus global neural context is proposed to achieve congruence with the situational context by convergence of spatial activity patterns in interacting networks to informationally consistent states. Currency with changes in situational context is maintained by ongoing disturbance and reestablishment of consistent states (Freeman 2006).

2 The Role of Neural Context in Cognition

Neural context at multiple scales

As a principle of brain function, neural context can be most easily demonstrated in relatively simple nervous systems, such as those of invertebrates. While these systems admittedly do not have the broad behavioral repertoire of primates, if contextual effects are indeed central to neural network operation, they should be present in simpler organisms. It has indeed been demonstrated in the Aplysia abdominal ganglion that the same neurons fire during performance of quite different behaviors (Wu et al., 1994). What appears to differentiate these behaviors is not the activity of a particular neuron, or group of neurons, but rather the overall activity patterns of an entire network. Such observations have been made in other invertebrate species across highly dissimilar behaviors (Popescu & Frost 2002), suggesting that the observed behavioral variation resides in the large-scale dynamics of entire networks rather than dedicated circuits (Kristan & Shaw 1997).

In the mammalian primary visual cortex (V1), neural context has been established as playing a major role in determining the receptive field properties of single neurons (Zipser et al., 1996; Gilbert 1998; Das & Gilbert 1999; Wörgötter & Eysel 2000; Gilbert et al., 2001; Stettler et al., 2002; Li et al., 2004). The receptive fields of V1 neurons were traditionally viewed as being spatially limited and tuned to simple stimulus attributes. However, it is now known that neural context influences the receptive field structure of V1 neurons. Thus, the response properties of these neurons are not determined solely by feedforward, convergent excitation from visual thalamic cells, but are also affected by lateral and feedback connections, some of which may be inhibitory. Neural contextual influences on V1 neurons reflect situational context in a variety of its forms. These include external situational factors derived from the global spatial and temporal characteristics of the visual scene. They also include internal situational factors such as the activation history of the local

cortical network in which the neuron is embedded, attentional influences, and the global state of arousal. The substantial body of evidence demonstrating these effects in V1 provides a vivid picture showing how situational context can affect neural operations through neural context.

Evidence that neural context also operates at a larger scale across cortical areas comes from studies showing that the processing of sensory input in one cortical area or region can depend on the processing status in another. Within the visual system, V1 neuron responses to illusory contours appear to depend on contextual guidance from V2 neurons that integrate spatial information over a broader spatial range (Lee 2003). Inter-sensory contextual modulation may also occur. For example, contextual modulation of responses in the visual cortex to visual stimuli can result from the concurrent processing of auditory stimuli (Bhattacharya et al. 2002).

As mentioned above, it has been hypothesized (Bressler 2004) that cortical context emerges from multiple recurrent interactions among cortical areas. An important prediction from this hypothesis is that the representation of categorical information in the cortex should be reflected by patterns of activity distributed across large cortical expanses rather than by the activity in a single specific area. Category-specificity has rapidly become a major focus in human neuroimaging research, exemplified by studies demonstrating face-category-specific responses in the fusiform gyrus (Kanwisher et al., 1997). However, a drawback of many such studies is that they employ very strict univariate statistical criteria that conceal all but the largest amplitudes in the activity patterns. Nonetheless, studies that have characterized the distributed response to faces have reported that greater category-specificity is revealed by the entire activity pattern in occipital and temporal cortices than by any specific area (Ishai et al., 1999; Haxby et al., 2001). Importantly, these studies have determined that the specificity of the distributed response is not dramatically altered if the regions typically associated with the category of interest are excluded.

The effect of neural context is seen in other cognitive operations as well. Working memory function, for example, is known to be supported by mutual influences among neurons in many different brain regions, and is not uniquely localized to prefrontal cortex (PFC) as was long thought (Fuster 1997, 2003). In a study of visual working memory, similar delay-period activity was observed in dorsolateral prefrontal and inferior temporal cortices, and cooling of either area induced similar effects on the activity in the other area (Fuster et al. 1985). Furthermore, a study of spatial working memory demonstrated nearly identical delay-period activity profiles in dorsolateral PFC and posterior parietal cortex (PPC) (Chafee & Goldman-Rakic 1998), and a follow-up study showed similar effects in either area from cooling of the other (Chafee & Goldman-Rakic 2000).

On balance, these studies support the conclusion that neural context operates in working memory through mutual interactions among distributed cortical association areas (Barash 2003). Regions such as dorsolateral PFC

and PPC seem central to working memory operations, but their contribution can only be realized by the particular set of interactions in which they engage at a given point in time.

Neural context and effective connectivity

The estimation of effective (or functional) connectivity provides strong evidence for variation in the recurrent interactions between neural elements that is thought to underlie neural contextual effects. To characterize interactions within the dorsal and ventral visual cortical processing streams (Ungerleider & Mishkin 1982) McIntosh et al. (1994) applied structural equation modeling to PET data to measure the effective connections specific to object (face matching) and spatial processing (location matching). Results from the right hemisphere analysis are presented in Fig. 1 (left hemisphere interactions did not differ between tasks). As expected, effects along the ventral pathway extending into the frontal lobe were stronger in the face-matching model, while interactions along the dorsal pathway to the frontal lobe were relatively stronger in the location-matching model.

Among posterior areas, the differences in path coefficients were mainly in magnitude. Occipitotemporal interactions between area 19v and area 37 were stronger in the face-matching model while the impact of area 17/18 to 19d and the occipitoparietal influences from area 19d to area 7 were stronger in the location-matching model. The model allowed for interactions between

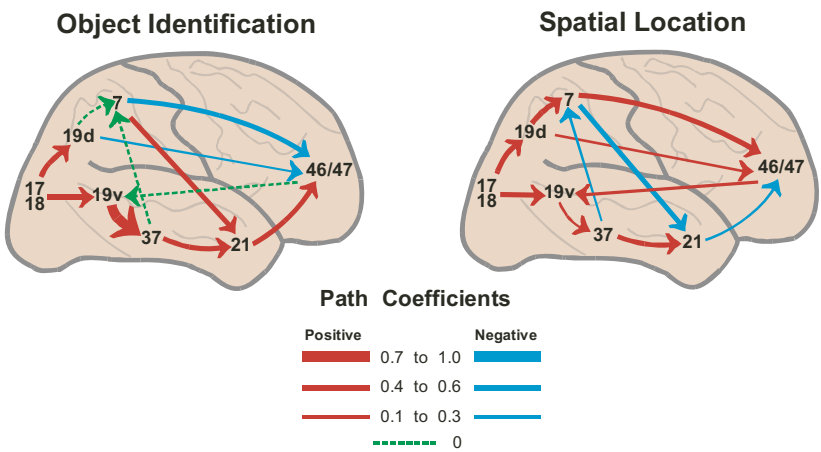


Fig. 1. Effective connectivity between cortical areas in the right hemisphere for object and spatial vision operations. The numbers on the cortical surface refer to Broadmann areas (d=dorsal, v=ventral). The arrows represent the anatomical connections between areas and the magnitude of the direct effect from one area to another is proportional to the arrow width for each path (Adapted from McIntosh et al. 1994)

the dorsal and ventral pathways with connections from area 37 to area 7 and from area 7 to area 21. In the right hemisphere, the interactions among these areas showed task-dependent differences in magnitude and sign. The temperoparietal influence of area 37 on area 7 was relatively stronger in the location-matching model. The parietotemporal influence of area 7 on area 21 showed a difference in sign between the two functional models. These results show that while the strongest positive interactions in each model may have been preferentially located within one or the other pathway, the pathways did not function independently, but exerted contextual modulatory influences on one another.

Another important result of this study is that, although the PFC did not show a difference in mean activity between tasks, processes involving PFC shifted depending on the task. The influence of the dorsal and ventral pathways on frontal cortex was similar in magnitude for the two tasks, but the origin of the positive and negative influences differed, implying that the qualitative nature of influence on the frontal lobe was different (positive influences in the location-matching model were from areas 7 and 19d, and in the face-matching model was from area 21). In terms of neural context, this result demonstrates that it is not an area's activity per se that is the key to understanding its contribution to a task, but rather its pattern of interaction with other areas in large-scale networks.

Network interactions that underlie cognitive operations are observable as differences in the effective connections between elements of the network. As illustrated above, if visual attention is directed to the features of an object, effective connections among ventral posterior cortical areas tend to be stronger, whereas visual attention directed to the spatial location of objects leads to stronger interactions among dorsal posterior areas. Another way that cognitive operations may be observed is through the modulation of effective connections that occurs when one area provides an enabling condition to foster communications between other areas. Such enabling effects may represent a primary mechanism whereby situational context is translated into neural context.

The most obvious example of neural context is top-down attentional control, whereby elements at higher processing levels can alter the processing mode of lower-level elements. In an fMRI study by Buchel and Friston (1997), subjects alternated between periods of overt attention to changes in a moving visual dot pattern and periods where they did not attend to the display. Two models were evaluated. In the first, a feedforward network from primary visual cortex (V1) to dorsal occipital cortex (V5) to PPC, visual attention was associated with elevated path coefficients as compared to inattention. The second model was an elaboration of the first, designed to assess whether the PFC had a modulatory influence on the effective connections between V5 and PPC. This second model, displayed in Fig. 2, revealed that PFC activity had a direct influence on PPC, as well as on the interaction term for the effect of V5 on PPC (PFC Mod $\rightarrow$ PP). This modulatory effect was shown to vary

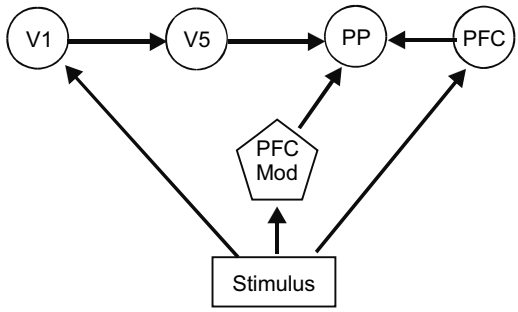


Fig. 2. Effective connectivity model for attentional modulation of visual processing of motion. Stimulus effects impact on V1 and PFC, and determine when attentional demands change via PFC Mod. The attentional effect was strongest at PFC, and through the modulatory effects, PFC also impacted the responsiveness of PP to the influence from V5. (Adapted from Buchel and Friston 1997)

in an activity-dependent manner, such that the effect from V5 to PPC was strongest when PFC activity was highest.

In a second study, Buchel et al. (1999) provided a convincing demonstration that changes in effective connectivity are directly related to learning. Training subjects to associate visually-presented objects with their location in space forced a learning-dependent change in the effective connections between dorsal and ventral visual processing streams. Furthermore, a remarkable correlation was found between the rate of learning and the change in the influence of the dorsal-stream PPC on the ventral-stream inferior temporal cortex.

A salient demonstration of neural context comes from a recent paper examining functional connectivity of the medial temporal lobe (MTL) in relation to learning and awareness (McIntosh et al., 2003). In a sensory learning paradigm, subjects were classified as AWARE or UNAWARE based on whether they noted that one of two tones predicted a visual event. Only AWARE subjects acquired and reversed a differential response to the tones, but both groups showed learned facilitation. The observation that MTL activity was related to learned facilitation in both groups at first appeared inconsistent with an earlier hypothesis that the MTL is critical for learning with awareness, but not when learning proceeds without awareness (Clark & Squire 1998). However, it was discovered that the functional connectivity patterns of the MTL were completely different for the two subject groups. In the AWARE group, dominant MTL effective connectivity was observed with prefrontal, occipital and temporal cortices, whereas in UNAWARE subjects, functional connectivity was more spatially restricted to inferior temporal cortex, thalamus and basal ganglia. We conclude that the MTL was involved in learning in both groups, but its functional role differed between the two groups because the neural context for its processing was different.

Another perspective on working memory emphasizes its close relation to sustained attention (McElree 2001; Fuster 2003; Deco & Rolls 2005; Bressler & Tognoli 2006). Both working memory and sustained attention involve activity in overlapping regions of PPC, PFC, and anterior cingulate cortex (ACC). In an fMRI study of the relationship between attention and working memory, Lenartowicz and McIntosh (2005) used two variants of a two-back working memory task: a standard version with strong attentional demands, and a cued version that more strongly promoted memory retrieval. Activation of ACC was found in both tasks, though it was more sustained in the standard condition. However, the regions functionally connected to the ACC, and the relation of the connectivity patterns to memory performance, differed completely between tasks. In the standard task, the observed pattern was related to a speed-accuracy tradeoff, with strong functional connection of ACC to PFC and PPC. In the cued task, the connectivity pattern was related only to better accuracy, and involved functional connections with middle and inferior PFC, and inferior temporal cortex. By virtue of these different patterns of functional connectivity, the contribution of ACC to attention- and memory-driven performance was similarly changed. In other words, although the activity of ACC was similar in both tasks, each task invoked a different neural context within which the ACC interacted, resulting in two very different behavioral profiles. The difference in neural context, and not in the activity of ACC per se, reflected the difference in the functional role that this region fulfilled.

In another study of ACC functional connectivity (Stephan et al., 2003), the question was examined of whether hemispheric functional asymmetry was determined by a word stimulus (short words, with one letter colored red) itself or by the task, i.e. the situational context. In one instance, subjects judged whether the word contained the letter "A", ignoring the red letter, and in another instance, they made a visuospatial judgment indicating whether the red letter was right or left of center. A direct comparison of the activity (measured with fMRI) revealed strong hemispheric differences. The letter task produced higher activity in the left hemisphere, while the visuospatial task produced higher activity in the right hemisphere. The ACC was similarly active in both tasks relative to baseline, but showed distinctly different patterns of effective connectivity between tasks. Specifically, during the letter task, the ACC was coupled to the left PFC; during the visuospatial task, the ACC was linked with the right PPC. These data are a compelling example of how situational context (in this case, task demands) can modulate the neural context within which a cortical area (i.e., the ACC) operates.

Disruption of contextual processing in cognitive disorders

We propose that the interplay between situational and neural context lies at the heart of normal brain operation. It follows that brain dysfunction should result from disruption of this interplay. In neurodegenerative diseases, neural

contextual effects may change as a result of primary damage, and also as the brain attempts to compensate for the degenerative process. In this regard, Grady et al. (2001) observed that patient with mild Alzheimer's Disease (AD), showed a distinctly different pattern of functional connectivity supporting working memory for faces, despite having similar behavioral profiles as age-matched controls. More direct evidence for new patterns of interactivity supporting cognition was provided by Grady et al. (2003), where increased functional connectivity in ventral prefrontal cortex was directly related to preserved memory performance in episodic memory of AD patients (Fig. 3).

In mental disorders, it is likely that the exact mapping between situational context and neural context is altered, such that changes in situational context are not properly reflected in neural context changes. In seeking to understand the neural basis of schizophrenia, some authors (Cohen et al. 1999)

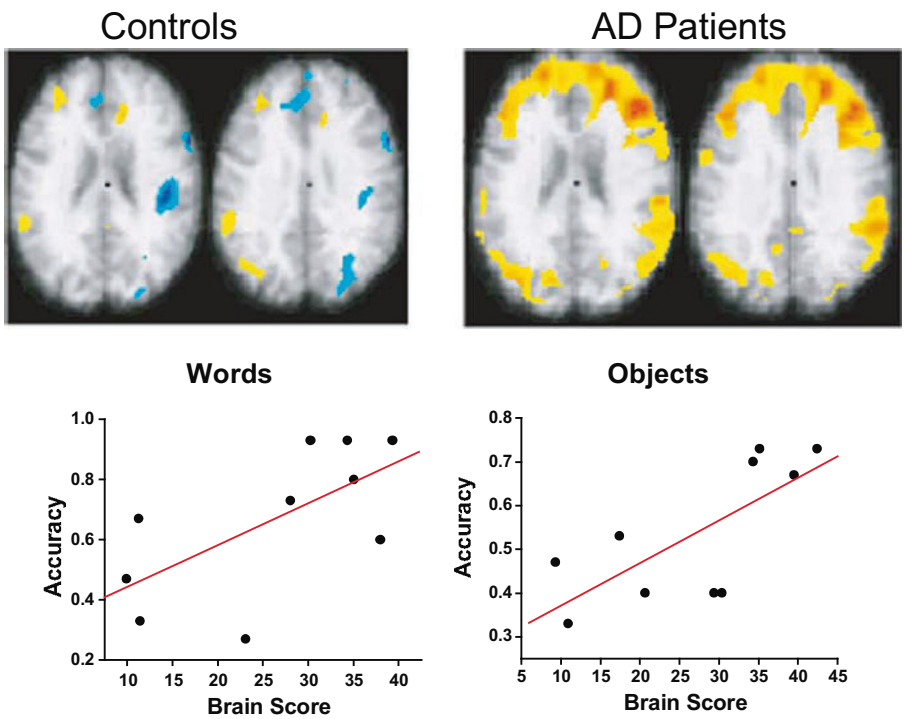


Fig. 3. Pattern of ventral prefrontal cortex (VPFC) functional connectivity in patients with mild Alzheimer's Disease (AD patients) and age-matched controls in an episodic memory task. Areas in yellow indicate strong positive functional connection (correlation) with VFPC, while blue indicates a negative correlation. Areas are plotted on an axial structural MRI (left is left). Scatterplots on the bottom indicate the relation to the functional connectivity pattern (brain scores) with memory accuracy in the AD patients, indicating the stronger the functional connection, the better the memory performance (Adapted from Grady et al. 2003)

have focused on the impaired utilization of situational context, whereas others (Bressler 2003b; Phillips & Silverstein 2003; Must et al., 2004; Dakin et al., 2005) have emphasized the impairment of neural contextual operations. Yet, these two aspects of contextual processing are likely to be related, as described above. The relation between situational and neural context may hold the key for both the understanding and treatment of mental disorders. Seminowicz et al. (2004), using estimation of effective connectivity, demonstrated distinctly different interaction patterns of limbic, cingulate and prefrontal regions across three groups of patients with major depression. Importantly, the groups were defined based on the form of therapy that was most effective in treating their depression. Patients responding to cognitive-behavioral therapy were distinguished from patients responding best to pharmacotherapy by the pattern of limbic-coritcal and cortico-cortical effective connections.

3 Concluding Remarks on the Generality of Neural Context

Although the notion of neural context may appear to be at odds with the idea of specialization of function in the brain, this is not the case. In our view, a brain area is only able to contribute to cognitive operations through the interactions that it undergoes with the other areas to which it is connected. From this perspective, a brain area plays a specialized role in any cognitive function by virtue of its unique position within the overall connectional framework of the brain. However, for the same brain area to be involved in a number of different functions does not necessarily imply that it exercises the same functional role in each. To the contrary, a large body of neuroimaging results indicates that different cognitive functions are associated with different neural contexts, and individual areas may contribute in a differentially specialized manner within each neural context. For example, since frontal cortical areas typically interact with parietal cortical areas in performing central executive functions (Collette & Van der Linden 2002), frontal or parietal areas may be considered as playing a specialized role in each function based on their contribution within the unique neural context associated with that function.

Clearly, we view neural context as a general effect that modulates the processing which occurs in any part of the brain. The neural context for processing in any area is firstly dependent on its connectivity with other brain areas, and secondly on the processes occurring in those other areas. Nonetheless, within the overall connectional architecture of the brain, some areas may occupy privileged positions for translating situational context into neural context.

As one example, the dorsolateral PFC may play a special role in establishing the neural context of working memory. Located at theapex of the frontal

executive hierarchy (Fuster 2003), this area is ideally situated to integrate information about situational context, such as the input modality and the type of response required in a given task situation. Our picture is that the PFC contributes to working memory by shaping the neural context of distributed sensory and motor networks through the modulatory influences that it exerts on them. Furthermore, its own neural context is shaped by the modulatory influences that it receives back from them. In this process, the PFC interacts with different other areas to instantiate different aspects of situational context, as when it interacts with the basal ganglia to maintain cross-temporal context (Dominey and Boussaoud 1997). By virtue of its different interactions, its own neural context is expected to depend on the situational context.

The hippocampus (along with surrounding medial temporal areas) also appears to occupy a privileged position with regard to the learning of situational contexts (Chun & Phelps 1999; Bucci et al., 2000; Smith & Mizumori 2006). In some theories of memory consolidation (e.g., Squire et al., 2004), the hippocampus acts gradually to strengthen the synaptic connections among neocortical areas representing the long-term memory content; after a sufficient degree of neocortical reorganization this memory content can be accessed independently of the hippocampus. From our perspective, this neocortical memory content is constrained by the global neocortical context that exists at the time that the consolidative processes are in effect. In this way, neural context may impact the integration of new memory into existing representations, and affect the ultimate accessibility of long-term memory to retrieval.

In this chapter, we have only touched on some of the main aspects of neural context with respect to the operation of large-scale neurocognitive networks. A number of additional facets of neural context are highly important in their bearing on neurocognitive function. The concept of contextual congruence, for example, may be a relevant aspect of that function. Laurienti et al. (2003) have demonstrated that the ACC and adjacent medial PFC are sensitive to the contextual congruence of multisensory input. Whether the congruence of this type of context, which is situational in nature, has a correspondence in some property of neural context, such as spatial coherence (Bressler 2004), remains to be tested.

This chapter has been concerned with the relation between neural context and one form of non-neural context, namely situational context. It is clear, however, that contextual effects can encompass a long list of other influences such as the personal and evolutionary history of the organism. We predict that the concept of neural context will take on increasing significance for brain research in coming years as researchers come to grips with the functional consequences of large-scale brain connectivity, and that it will come to be seen as a common form of implementation for a number of different types of non-neural context.

Acknowledgments

Supported by grant MH64204 from the National Institute of Mental Health to SLB and Canadian Institute of Health Research and JS McDonnell grants to ARM.

References

Andras P (2005) Pattern computation in neural communication systems. Biol Cybern 92, 452–460

Bar M (2004) Visual objects in context. Nat Rev Neurosci 5, 617–629

Barash S (2003) Paradoxical activities: insight into the relationship of parietal and prefrontal cortices. Trends Neurosci 26, 582–589

Beck DM, Kastner S (2005) Stimulus context modulates competition in human extrastriate cortex. Nat Neurosci 8, 1110–1116

Becker S, Hinton GE (1992) Self-organizing neural network that discovers surfaces in random-dot stereograms. Nature 355, 161–163

Beggs JM, Klukas J, Chen W (2007) Connectivity and dynamics in local cortical networks. This Volume

Bhattacharya J, Shams L, Shimojo S (2002) Sound-induced illusory flash perception: role of gamma band responses. Neuroreport 13, 1727–1730

Breakspear M, Jirsa V (2007) Neuronal dynamics and brain connectivity. This Volume

Bressler SL (1995) Large-scale cortical networks and cognition. Brain Res Rev 20, 288–304

Bressler SL (2002) Understanding cognition through large-scale cortical networks. Curr Dir Psychol Sci 11, 58–61

Bressler SL (2003a) Context rules. Commentary on Phillips WA & Silverstein SM, Convergence of biological and psychological perspectives on cognitive coordination in schizophrenia. Behav Brain Sci 26, 85

Bressler SL (2003b) Cortical coordination dynamics and the disorganization syndrome in schizophrenia. Neuropsychopharmacology 28, S35–39

Bressler SL (2004) Inferential constraint sets in the organization of visual expectation. Neuroinformatics 2, 227–238

Bressler SL, Tognoli E (2006) Operational principles of neurocognitive networks. Int J Psychophysiol 60, 139–148

Brovelli A, Ding M, Ledberg A, Chen Y, Nakamura R, Bressler SL (2004) Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. Proc Natl Acad Sci USA 101, 9849–9854

Bucci DJ, Phillips RG, Burwell RD (2000) Contributions of postrhinal and perirhinal cortex to contextual information processing. Behav Neurosci 114, 882–894

Buchel C, Friston KJ (1997) Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cereb Cortex 7, 768–778

Buchel C, Coull JT, Friston KJ (1999) The predictive value of changes in effective connectivity for human learning. Science 283, 1538–1541

Chafee MV, Goldman-Rakic PS (1998) Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. J Neurophysiol 79, 2919–2940

Chafee MV, Goldman-Rakic PS (2000) Inactivation of parietal and prefrontal cortex reveals interdependence of neural activity during memory-guided saccades. J Neurophysiol 83, 1550–1566

Chun MM (2000) Contextual cueing of visual attention. Trends Cogn Sci 4, 170–178

Chun MM, Phelps EA (1999) Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. Nat Neurosci 2, 844–847

Clark RE, Squire LR (1998) Classical conditioning and brain systems: the role of awareness. Science 280, 77–81

Cohen JD, Barch DM, Carter C, Servan-Schreiber D (1999) Context-processing deficits in schizophrenia: converging evidence from three theoretically motivated cognitive tasks. J Abnorm Psychol 108, 120–133

Collette F, Van der Linden M (2002) Brain imaging of the central executive component of working memory. Neurosci Biobehav Rev 26, 105–125

Dakin S, Carlin P, Hemsley D (2005) Weak suppression of visual context in chronic schizophrenia. Curr Biol 15, R822–824

Das A, Gilbert CD (1999) Topography of contextual modulations mediated by short-range interactions in primary visual cortex. Nature 399, 655–661

Deco G, Rolls ET (2005) Attention, short-term memory, and action selection: a unifying theory. Prog Neurobiol 76, 236–256

DeFelipe J, Alonso-Nanclares L, Arellano JI (2002) Microstructure of the neocortex: comparative aspects. J Neurocytol 31, 299–316

Dominey PF, Boussaoud D (1997) Encoding behavioral context in recurrent networks of the fronto-striatal system: a simulation study. Cogn Brain Res 6, 53–65

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1, 1–47

Finger S (1994) The Origins of Neuroscience: A History of Explorations into Brain Function. Oxford University Press, New York

Freeman WJ (2003) The wave packet: an action potential for the 21st century. J Integr Neurosci 2, 3–30

Freeman WJ (2006) A cinematographic hypothesis of cortical dynamics in perception. Int J Psychophysiol 60, 149–161

Friston KJ, Price CJ (2001) Dynamic representations and generative models of brain function. Brain Res Bull 54, 275–285

Fuster JM (1997) Network memory. Trends Neurosci 20, 451–459

Fuster JM (2003) Cortex and Mind: Unifying Cognition. Oxford University Press, New York

Fuster JM, Bauer RH, Jervey JP (1985) Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. Brain Res 330, 299–307

Gilbert CD (1998) Adult cortical dynamics. Physiol Rev 78, 467–485

Gilbert CD, Sigman M, Crist RE (2001) The neural basis of perceptual learning. Neuron 31, 681–697

Grady CL, Furey ML, Pietrini P, Horwitz B, Rapoport SI (2001) Altered brain functional connectivity and impaired short-term memory in Alzheimer's disease. Brain 124, 739–756

Grady CL, McIntosh AR, Beig S, Keightley ML, Burian H, Black SE (2003) Evidence from functional neuroimaging of a compensatory prefrontal network in Alzheimer's disease. J Neurosci 23, 986–993

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425–2430

Hepp-Reymond M, Kirkpatrick-Tanner M, Gabernet L, Qi HX, Weber B (1999) Context-dependent force coding in motor and premotor cortical areas. Exp Brain Res 128, 123–133

Herzog MH, Fahle M (2002) Effects of grouping in contextual modulation. Nature 415, 433–436

Ishai A, Ungerleider LG, Martin A, Schouten JL, Haxby JV (1999) Distributed representation of objects in the human ventral visual pathway. Proc Natl Acad Sci USA 96, 9379–9384

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17, 4302–4311

Kay J, Phillips WA (1997) Activation functions, computational goals and learning rules for local processors with contextual guidance. Neural Comput 9, 895–910

Kristan WB Jr, Shaw BK (1997) Population coding and behavioral choice. Curr Opin Neurobiol 7, 826–831

Laurienti PJ, Wallace MT, Maldjian JA, Susi CM, Stein BE, Burdette JH (2003) Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. Hum Brain Mapp 19, 213–223

Lee TS (2003) Computations in the early visual cortex. J Physiol Paris 97, 121–139

Lenartowicz A, McIntosh AR (2005) The role of anterior cingulate cortex in working memory is shaped by functional connectivity. J Cogn Neurosci 17, 1026–1042

Li W, Piech V, Gilbert CD (2004) Perceptual learning and top-down influences in primary visual cortex. Nat Neurosci 7, 651–657

Lopes da Silva FH, Pijn JP, Wadman WJ (1994) Dynamics of local neuronal networks: control parameters and state bifurcations in epileptogenesis. Prog Brain Res 102, 359–370

McElree B (2001) Working memory and focal attention. J Exp Psychol Learn Mem Cogn 27, 817–835

McIntosh AR (1999) Mapping cognition to the brain through neural interactions. Memory 7, 523–548

McIntosh AR, Grady CL, Ungerleider LG, Haxby JV, Rapoport SI, Horwitz B (1994) Network analysis of cortical visual pathways mapped with PET. J Neurosci 14, 655–666

McIntosh AR, Rajah MN, Lobaugh NJ (2003) Functional connectivity of the medial temporal lobe relates to learning and awareness. J Neurosci 23, 6520–6528

Must A, Janka Z, Benedek G, Keri S (2004) Reduced facilitation effect of collinear flankers on contrast detection reveals impaired lateral connectivity in the visual cortex of schizophrenia patients. Neurosci Lett 357, 131–134

Passingham RE, Stephan KE, Kotter R (2002) The anatomical basis of functional localization in the cortex. Nat Rev Neurosci 3, 606–616

Phillips WA, Silverstein SM (2003) Convergence of biological and psychological perspectives on cognitive coordination in schizophrenia. Behav Brain Sci 26, 65–82

Phillips WA, Singer W (1997) In search of common foundations for cortical computation. Behav Brain Sci 20, 657–722

Phillips WA, Floreano D, Kay J (1998) Contextually guided unsupervised learning using local multivariate binary processors. Neural Netw 11, 117–140

Popescu IR, Frost WN (2002) Highly dissimilar behaviors mediated by a multifunctional network in the marine mollusk Tritonia diomedea. J Neurosci 22, 1985–1993

Seminowicz DA, Mayberg HS, McIntosh AR, Goldapple K, Kennedy S, Segal Z, Rafi-Tari S (2004) Limbic-frontal circuitry in major depression: a path modeling metanalysis. Neuroimage 22, 409–418.

Siegel RM, Read HL (2001) Deterministic dynamics emerging from a cortical functional architecture. Neural Netw 14, 697–713

Smith DM, Mizumori SJ (2006) Learning-related development of context-specific neuronal responses to places and events: the hippocampal role in context processing. J Neurosci 26, 3154–3163

Sporns O, Tononi G (2007) Structural determinants of functional brain dynamics. This Volume

Sporns O, Tononi G, Edelman GM (2007) Connectivity and complexity: the relationship between neuroanatomy and brain dynamics. Neural Netw 13, 909–922

Squire LR, Stark CE, Clark RE (2004) The medial temporal lobe. Annu Rev Neurosci 27, 279–306

Stettler DD, Das A, Bennett J, Gilbert CD (2002) Lateral connectivity and contextual interactions in macaque primary visual cortex. Neuron 36, 739–750

Thagard P, Verbeurgt K (1998) Coherence as constraint satisfaction. Cognit Sci 22, 1–24

Tononi G, Sporns O, Edelman GM (1992) Reentry and the problem of integrating multiple cortical areas: simulation of dynamic integration in the visual system. Cereb Cortex 2, 310–335

Ungerleider LG, Mishkin M (1982) Two cortical visual systems. In DJ Ingle, MA Goodale, RJW Mansfield (Eds.), Analysis of Visual Behavior 549–586. MIT Press, Cambridge, MA

Wörgötter F, Eysel UT (2000) Context, state and the receptive fields of striatal cortex cells. Trends Neurosci 23, 497–503

Wu JY, Cohen LB, Falk CX (1994) Neuronal activity during different behaviors in Aplysia: a distributed organization? Science 263, 820–823

Zipser K, Lamme VA, Schiller PH (1996) Contextual modulation in primary visual cortex. J Neurosci 16, 7376–7389

Neural Coordination Dynamics of Human Sensorimotor Behavior: A Review

Kelly J Jantzen and JA Scott Kelso

Human Brain and Behavior Laboratory
Center for Complex Systems and Brain Sciences, Florida Atlantic University,
Boca Raton, FL, 33431

1 Introduction

Coordination symbolizes one of the most fundamental, yet arguably least understood ability of living things. Humans effortlessly coordinate complex interactions between their bodies and the world in order to accomplish even the simplest of tasks. Coordination Dynamics, defined broadly as the science of coordination (see contributions, e.g. in Tschacher and Dauwalder 2003; Jirsa and Kelso 2004) describes, explains and predicts how patterns of coordination form, adapt, persist and change in natural systems. It uses the language, methods and tools of self-organizing dynamical systems (e.g. Haken 1983; Nicolis and Prigogine 1989) to provide a conceptual and theoretical framework for understanding coordination at multiple levels, from behavior to brain (Kelso 1995; Haken 1996; Kelso et al. 1999). A primary goal of coordination dynamics is to understand the laws, principles and mechanisms governing how patterns of behavior form in space and time; how these patterns are maintained; how they change in response to environmental or physiological demands; and how they are reorganized in an adaptive way, often allowing for the adoption of new patterns. "Understanding" means identifying both the essential variables and parameters that characterize dynamic patterns of coordination on a given level of description and the usually low-dimensional pattern dynamics that gives rise to a broad range of complex, dynamic behaviors. As a science of coordination that embraces both structure and function in living things, coordination dynamics deals essentially with *informationally* coupled self-organizing systems. That is, in coordination dynamics otherwise independent component subsystems are coupled by functionally meaningful information.

In studies of coordination dynamics, behavioral coordination has proven a fertile entry point into uncovering principles and mechanisms of human action (Kelso 1995; Jirsa and Kelso 2004). Within this conceptual and theoretical framework, coordination is defined in terms of collective or coordination variables that specify the spatiotemporal ordering between component

parts. In the vicinity of critical points, emergent behavior has been shown to be governed by the low-dimensional dynamics of these collective variables (e.g., Haken, Kelso & Bunz, 1985). This low-dimensional dynamics is revealed by manipulating one or more control parameters whose role is simply to move the system through a series of state changes. Adopting this strategy in the case of human behavioral coordination has revealed that the relative phase between component subsystems is a crucial collective variable and that the rate of coordination is a key control parameter (Kelso 1984; Haken et al. 1985). Thus, when rate is systematically increased, instabilities in coordination develop and spontaneous switches between patterns of coordination ensue. It is these transitions and the features that surround them that have established the self-organizing nature of human behavior (see Kelso, et al., 1987; Schöner and Kelso 1988 for early reviews).

A deeper understanding of human behavior and cognition rests on uncovering how the rich dynamics of human behavioral pattern formation is rooted in underlying brain processes. In this chapter we take the case of sensorimotor coordination as an entry point for determining the relation between the dynamics of human behavior and the dynamics of brain function. We begin with a brief description of levels of observation, including the kinds of measures that may help us to connect behavioral and neural levels (see also Fuchs et al. 1999; Jirsa et al. 1999; Kelso et al. 1999). We then go on to describe the experimental 'workhorse' paradigm used to study human behavioral pattern formation and the basic phenomena it reveals. This is followed by a brief overview of the theoretical and mathematical framework employed to explain these phenomena. Such concepts aid in identifying the key variables that govern coordination and hence provide experimental insight into its nature. Having identified key coordination variables, we then review the literature that has either directly or indirectly shown how patterns of human sensorimotor behavior may be mapped onto neural function. Finally, we conclude with some key questions that future research needs to address and how technological and analytical advances in functional neuroimaging may aid in finding the answers.

2 Levels of Observation

At the behavioral level we take as our entry point sensorimotor or perceptual motor coordination (Kelso et al. 1990). This can be broadly defined as coordinating one's actions in space and time with environmental stimuli. Such stimuli can vary widely in scope ranging from a simple rhythmic stimulus presented over one or more perceptual modalities to multifaceted biologically relevant stimuli such as the actions of another person (e.g., DeGuzman, et al., 2005; Oullier, et al., 2003). Regardless, the idea is to identify key measures of the coordination between interacting elements and the dynamics governing this interaction. Much research over the last two decades has demonstrated that

relative phase is a relevant collective or coordination variable that: a) captures (multiple) patterns of behavior of coupled systems; b) reveals specific features of the underlying dynamics of the coordinated system such as loss of stability; and c) demonstrates the existence of pattern switching. Operationally, relative phase is defined as the time difference between an environmental event (e.g., a simple tone beep, tactile vibration or visual flash), and the associated behavioral action (e.g., a simple finger flexion) divided by the stimulus period. Since we are dealing in the first instance with rhythmic behavior, this normalized time difference is often expressed in radians or degrees by multiplying it by 2π or 360 respectively (see Jirsa & Kelso, 2005 for a more generic treatment).

At the neural level we concentrate on relatively macroscopic measures of brain function derived from electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI) and positron emission tomography (PET). These technologies supply a variety of measures of the activity of neural ensembles that, when combined with various analytic methods provide insight into the large scale functioning of the human brain. It should be noted that each of these recording tools provides unique information about large-scale patterns of neural activity and possesses advantages and disadvantages relative to one another. The latter will become clear as we proceed through forthcoming sections.

Current technologies restrict us to questions concerning how the rich dynamics of behavioral coordination originates from activity arising at the level of macroscopic brain function. Of course, this is not to say that other scales of observation are uninformative and do not provide additional insight into the questions addressed here. Neural activity unfolds over multiple spatial and temporal scales (e.g. cortical column, local field potential, single unit, ion channels, neurotransmitters, gene expression, etc.) that are likely to be important to understanding coordination at other levels. As for sensorimotor behavior, work at intermediate levels of observation suggests that biomechanical factors can play a role in modulating the intrinsic dynamics of human behavioral coordination (e.g. Ridderikhoff et al. 2004). For example, there appears to be a preference for recruitment of specific muscular synergies during coordination as revealed by the observation of transitions from extension on the beat to flexion on the beat (Carson et al. 1996; but see also Kelso, et al., 2001). This finding has been couched in terms of the modulation of central supraspinal mechanisms on spinal tendencies favoring the entrainment of flexors. Similarly, in wrist supination/pronation the most stable pattern (pronation or supination on the beat) is determined by the location of the axis of rotation with respect to the skeletal structure (Byblow et al. 1995; Carson and Riek 2000) suggesting that altering musculoskeletal constraints impacts the stability of performed coordination patterns. In short, although we concentrate here on understanding relative phase dynamics in terms of large scale neural activity we also recognize that the rich repertoire of both

bimanual (Swinnen 2002) and unimanual coordination (Ridderikhoff et al. 2004) arises from a coalition of constraints (Kelso et al. 2001; Temprado and Swinnen 2005) that includes, but is not limited to the combined influence of supraspinal, musculoskeletal and environmental factors.

3 Behavioral Level: Stability and Instability

To illustrate the key ideas and features of coordination dynamics, we describe a simple experimental system that treats the problem of sensorimotor coordination as a pattern forming process (Kelso et al. 1990). In the paradigmatic case, the temporal coupling is between a finger flexion/extension movement and a periodic auditory beep. The task requires participants to coordinate in one of two or more possible patterns. For the synchronized patterns movements are coordinated such that peak flexion of the movement occurs coincident with the metronome beat. Syncopation is accomplished by placing peak flexion directly in between two consecutive beats. In Kelso et al. (1990) the metronome was systematically increased in frequency from 1 Hz to 3.5 Hz in steps of 0.25 Hz. Instructions emphasized the importance of maintaining a 1:1 relationship between finger movement and metronome at all times, and if the pattern changed "not to intervene" or to intentionally restore the pattern (see also Kelso et al. 1988; Scholz and Kelso 1990). When finger flexion and metronome occur simultaneously in time the relative phase difference between them is 0°. Perfect syncopation is achieved when the relative phase difference is 180°.

A rich and reproducible set of behaviors flows from this paradigm. At low metronome rates a strong tendency exists toward frequency- and phase-locking in both conditions. That is, both patterns are performed in a stable manner with relative phase between finger and metronome close to 0 or 180 degrees. As the rate of the movement is parametrically increased, a spontaneous transition from syncopation to synchronization is observed at a critical frequency marking the change from a bistable regime (where both patterns are possible and may be said to coexist) to a monostable regime where synchronization predominates. In some cases further increases in rate result in the loss of entrainment between the finger and metronome resulting in-phase wandering and various patterns of phase drift. Transitions between coordination patterns are typically preceded by an increase in the variability of the relative phase between sensory and motor events. Such fluctuation enhancement in the vicinity of transition points indicates a characteristic loss of stability typical of nonequilibrium pattern forming systems in nature (Haken, 1983). Indeed, the foregoing experiments and resultant theoretical modeling (Kelso, et al., 1990) have been taken to provide strong evidence that action-perception follows laws of general pattern forming processes (Turvey, 2004). The symmetry breaking property of the theory developed in the Kelso et al (1990) work has proven to be generalizable across different experimental paradigms, thereby providing a framework for understanding and investigating human behavior

in a variety of contexts. For example, the same dynamical features such as loss of stability, critical slowing down and qualitative changes in the order parameter have been observed across different task contexts such as coordination between homologous limbs (Kelso 1984), non-homologous limbs (Kelso and Jeka 1992; Jeka and Kelso 1995; Swinnen et al. 1997), between persons and their environment (Kelso et al. 1990; Jirsa et al. 2000) and even in social coordination between people (Schmidt et al. 1998; 1999; Oullier, et al., 2003; Richardson et al., 2005). Similar principles govern perceptual grouping and coordination as nicely illustrated by the classic bistable properties of reversible figures such as the Necker cube (e.g. Kelso, et al., 1995), the visual perception of spatiotemporal in-phase and antiphase moving stimuli (Hock et al. 1993; Bingham et al. 1999; Zaal et al., 2000) and speech categorization (Tuller et al. 1994; Case et al. 1995). For present purposes, the key aspect of this essentially 'nonlinear paradigm' is that it provides a convenient entry point for investigating the dynamics of pattern formation and change at both behavioral and brain levels.

4 Theoretical Foundations

The theoretical model first proposed by Haken, Kelso and Bunz (1985), and its fundamental extensions to include stochastic (Schöner et al. 1986) and symmetry breaking features (Kelso et al. 1990) form the foundation for the science of coordination and provide a basis for understanding the self-organizing dynamics of human behavioral and neural function (Kelso 1995; Haken 1996; Bressler and Kelso 2001). The original HKB formulation assumed, for simplicity's sake, identical (homogeneous) elements. Here we provide a brief introduction to the computational model that accounts for the basic coordination phenomena in terms of the non-linear coupling between heterogeneous coordinating elements (Kelso et al. 1990). In this theoretical model, the coordination between finger flexion/extension movements and the environment is captured by the relative phase (ϕ) and represented by a nonlinear equation of motion that specifies the relative phase dynamics:

$$\dot{\phi} = \delta\omega - a\sin\phi - b\sin 2\phi + \sqrt{Q}\varepsilon t$$

The parameter $\delta\omega$ accounts for inevitable differences between the individual coordinating elements, here different intrinsic oscillatory frequencies. The parameters a and b govern the strength of the coupling, the ratio (b/a) corresponding to a control parameter of rate or frequency in experiments. Thus, a decrease in the ratio signifies a decrease in the period or duration of coupled dynamical behavior. The final parameter ε_t represents the presence of (delta correlated) noise fluctuations of strength Q.

For high values of the coupling ratio b/a (Fig. 1; left) the system exhibits bistability: two modes of behavioral coordination coexist for the same parameter set. Stable states of coordination, indicated by points where the negative

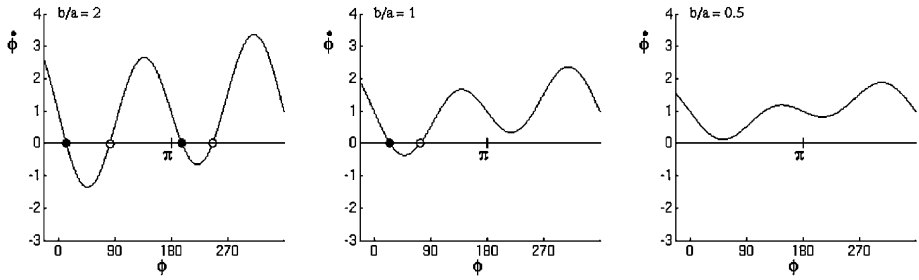


Fig. 1. The theoretical model of Kelso et al. (1990) showing plots of the relative phase dynamics under conditions of high (left), medium (middle) and low (right) coupling strength with $\delta\omega$ set at a small value. As a result of symmetry breaking, notice the fixed points of relative phase are not exactly at 0 and 180 deg. On the left, the coordinative system is in a bistable regime where both in-phase and antiphase coordination states exist (filled circles). Decreasing the coupling results in a switch to a monostable state where only in-phase coordination is supported (middle). Further decreases in coupling, in combination the symmetry breaking term $\delta\omega$, abolish all stable states leaving only the remnants of attractors in the landscape or phase portrait

slope of the function crosses the $\dot{\phi}$ axis (solid circles; open circles indicate unstable states) occur at relative phase relationships near 0 and 180 degrees. These values of the coordination or collective variable correspond to patterns of synchronization and syncopation respectively. As the ratio is decreased the system reaches a critical point and undergoes a switch from a bistable to a monostable regime where only the synchronized coordination pattern is stable. Note in the middle panel of Fig. 1 that only a single fixed point exists near 0 degrees and the negative going portion near syncopation no longer crosses the $\dot{\phi}$ axis. Further decreases in b/a result in destabilization of the in-phase pattern corresponding to a complete lack of stable and unstable fixed points such that the function is lifted entirely off the axis (Fig. 1, right).

The fact that the function retains its curvature is taken as evidence for remnants of attraction toward previously stable states. This effect is due to broken symmetry that arises as a result of intrinsic differences in the individual coordinating elements (embodied by the $\delta\omega$ parameter). Such *metastability* represents a more flexible form of self-organization where the component parts are free to express their individual behavior though still working together in loose association (Kelso, 1995; Friston, 1997). Systematically guiding the system through an appropriate parameter range exposes qualitatively different behavioral regimes and provides an entry point for studying neural mechanisms of pattern selection, formation and change—a kind of dynamical decision-making. Some of the key questions that arise from this theoretical and experimental programme that aims to connect the dynamics of behavioral coordination with the dynamics of neural activity are:

1. How are intrinsic differences in stability between syncopated and synchronized patterns of sensorimotor coordination manifest at the level of brain structure and function?
2. Since increases in movement rate for the syncopated pattern result in a decrease in coordinative stability whereas comparable decreases are not observed for synchronization, how are such changes in rate and stability manifest at the neural level?
3. Typical of informationally-based self-organizing dynamical systems (aka coordination dynamics), at a critical value of the control parameter of rate or frequency, spontaneous switches in pattern from synchronization to syncopation occur. Is a similar reorganization observed at the level of the brain? If so, what is the neuroanatomy and neurophysiology behind such pattern switching? *Putting our cards on the deck, we wish to emphasize that although switching is common in the self-organizing behavioral and neural systems described by coordination dynamics, this by no means implies the existence of "switches"(cf. Martin, 2006).*

In the following we review the imaging literature relevant to establishing connections between brain and behavioral dynamics in light of the foregoing questions. Because different technological and analytical approaches provide substantially different information about aspects of cortical and sub-cortical structure ∼ function the remainder of this review is organized according to the methods used in acquiring or analyzing information from the brain.

5 Evoked Responses and the Search for Shared Dynamics

Early studies in the present framework employed sensorimotor coordination tasks as a means to uncover the link between the dynamics of behavior and the dynamics of the brain; connecting these levels by virtue of their shared dynamical properties (Fuchs et al. 1992; Kelso, 1981; Kelso et al. 1991; 1992). The high temporal resolution of electroencephalography and magnetoencephalography was exploited to quantify the relationship between behavioral patterns and spatiotemporal patterns of neural activity. Common features of the dynamics expressed at both levels of description were taken as evidence that similar principles of (self) organization govern pattern formation in brain and behavior. Of particular initial interest was the identification of qualitative changes in the pattern of neural activity that occurred simultaneously with transitions between coordination patterns.

In the first of these studies, a 37 channel SQUID (Superconducting Quantum Interference Device) sensor array was used to measure the spatiotemporal patterning of neuromagnetic fields generated by the intracellular dendritic current flow of large ensembles of cortical neurons located over left temporoparietal cortex (2 cm. posterior to C3) during auditory-motor coordination

(viz. Kelso et al. 1990). Unlike EEG signals that are subject to blurring at the various tissue boundaries, magnetic signals pass unimpeded through skull, scalp and cerebrospinal fluid providing a relatively undistorted measure of the neural sources underlying coordination patterns and pattern switching. The task was to syncopate right index finger flexion in time with an auditory stimulus that parametrically increased in rate from 1.0 Hz to 3.25 Hz in 0.25 Hz increments. Under instructions to maintain a 1:1 stimulus-response coordination, increases in rate resulted in a gradual loss of stability of and a transition from syncopation to synchronization at a critical frequency of 1.75–2.00 Hz.

This neurophysiological data may be summarized by three main findings. First, time-frequency analysis showed a frequency-dependent change in the temporal relationship between rate and the dominant frequency in the averaged MEG signal. At slow, pre-transition plateaus (< 1.75 Hz) the frequency of the neural response was at the fundamental frequency of the stimulus/response. Following the transition, the spectral pattern's dominant Fourier component changed to twice the stimulus frequency (Fuchs et al. 1992), a finding interpreted in terms of a period doubling bifurcation in forced nonlinear systems (see Kelso & Fuchs, 1995 for a possible model). The second finding was a qualitative change in the phase relations of the MEG sensors occurring coincident in time with the period doubling and with the behavioral transition. Relative phase was employed to capture, within a single quantity, both the spatiotemporal organization between the brain signals and the auditory stimulus as well as the finger position and the auditory stimulus. Particularly the anterior sensors demonstrated a phase shift of π at the point of the behavioral transition signifying a similar transition in relative phase for both brain and behavior (Kelso et al., 1991; 1992). Third, using principal component analysis, it was determined that a majority of the variance (97%) in the spatiotemporal signal could be accounted for by only four temporally evolving spatial patterns (Fuchs et al. 1992; Kelso et al. 1992). Theoretical modeling used coupled differential equations to capture the observed facts (Jirsa et al. 1994), two of the mode amplitudes acting as collective variables or order parameters to which lower amplitudes are "enslaved" (in the language of synergetics). These spatial modes demonstrated notable qualitative effects in the time and frequency domain in relation to the occurrence of the behavioral transition. Unfortunately, due to the limited coverage of the MEG sensor array, specific neuroanatomical substrates could not be determined. However, given the frontal concentration of activity, involvement of pre-motor and prefrontal regions was assumed (Kelso et al. 1992). Nevertheless, the key point is that when a complex system like the brain is placed in a behavioral context that requires pattern formation and switching, it appears to be governed by just a few order parameters. This supports the general view that the brain operates (or can be induced to operate) close to instability where only a few key quantities capture its behavior and the system is governed by low-dimensional dynamics (Haken, 1996; Kelso, 1992; 1995). Significantly, it was possible to

derive these equations from a more microscopic approach in which neurons are defined in terms of their pulse rates and dendritic currents (Jirsa and Haken 1997; Jirsa et al. 1998). The latter of course are very relevant for the present discussion, since they give rise to the magnetic fields observed by MEG.

Subsequent work confirmed some of the findings in the Kelso et al. (1991; Kelso et al. 1992) MEG experiment while calling others into question. Daffersthofer and colleagues provided a replication of the original experiments using a 151 sensor MEG system that provided extended coverage and 36 EEG electrodes located primarily over bilateral motor areas (Daffertshofer et al. 2000). In line with the earlier studies, a phase shift of approximately π was observed simultaneously in behavior, EEG and MEG. This phase shift was broadly distributed across the MEG sensors suggesting a global reorganization of neural activity associated with the behavioral transition. In contrast to the analysis by Fuchs et al. (1992), no post-transition period doubling was observed leading to a suggestion that such a finding may have resulted from the specific method of analysis employed. Several arguments were subsequently made against this conclusion (Fuchs et al. 2000a) with a second MEG study definitively revealing the source of the observed period doubling (Fuchs et al. 2000b).

Further insight came in the form of EEG studies that provided homogeneous full head coverage to investigate spatial and temporal properties of activity arising from the entire cortical surface during both synchronized and syncopated coordination (Wallenstein et al. 1995; Mayville et al. 1999). The inclusion of synchronization is important to control for the effect of movement frequency since synchronization does not change in stability or switch pattern when performed at the same rates as syncopation (Kelso et al. 1990). Large-scale reorganization of the dominant pattern of neural activity over contralateral sensorimotor cortex was observed coincident with transitions in coordination at the behavioral level (Wallenstein et al. 1995; Mayville et al. 1999). Importantly, this reorganization was associated with syncopation and not synchronization indicating that the reorganization was related to changes in pattern stability and not simply a product of changes in movement rate *per se*. Experimentally induced effects were confined to electrodes overlying contralateral central and parietal regions suggesting that dynamic changes in cortical function associated with behavioral change occurred principally in primary sensorimotor and premotor cortex.

In a definitive study, Fuchs and colleagues (Fuchs et al. 2000b) proposed an explanation for previous findings in terms of the interplay between auditory and motor evoked signals. In this work the syncopation paradigm was investigated using a 143 channel SQUID array allowing for the measurement of neuromagnetic signals from the entire cortex. As in previous studies, movements began at 1.0 Hz in a syncopated pattern of coordination and increased in rate by 0.25 Hz after every 10 cycles to a maximum of 2.75 Hz. Auditory (listening to tones presented with a randomized ISI from 2–4 seconds) and

motor alone (self-paced movements separated by intervals of approximately 3s) conditions acted as controls.

Based on the spatiotemporal decomposition of the motor related response, the authors concluded that the π phase shift was due to the shifting in time of the motor response with respect to the auditory metronome. A shift of this kind reflects the tight time and velocity coupling between flexion of the index finger and the evoked magnetic response (Kelso et al. 1998). Comparison between the topographic patterns observed during coordination and those generated during control conditions indicated that at low movement rates the pattern of neural activity appeared to be dominated by auditory related activity while at higher rates a motor pattern was most prevalent. Principal component analysis was again used to decompose the signal into dominant spatial patterns and their temporal evolution. During pre-transition plateaus a majority of the variance was accounted for by a spatial mode that closely resembled the auditory control. Post-transition plateaus, on the other hand, were dominated by a clearly motor related mode. Projecting the data from each plateau onto a basis composed of the auditory- and motor-only controls demonstrated that the auditory related pattern (characterized by the N1m response) decreased in amplitude with increasing frequency while the motor component remained stable throughout (Fig. 2). Moreover, the period doubling reported previously (Fuchs et al. 1992) was shown to result directly from this interaction. Thus, at low movement rates the primarily monophasic auditory evoked response dominates the power spectrum resulting in a dominant

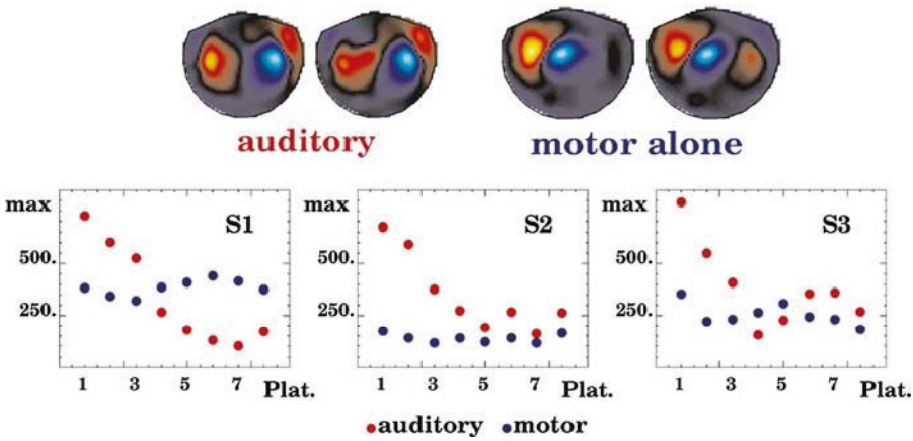


Fig. 2. Shown are the relative strengths of the spatial modes corresponding to the auditory (red, top left) and motor (blue, top right) components of the sensorimotor task. As movement rate increases (left to right in bottom panels) the strength of the auditory component systematically decreases while the motor component remains fairly constant for all rates. The place where they overlap corresponds approximately to the behavioral transition. The three panels show data from different subjects. Adapted from Fuchs et al. (2000)

frequency at the stimulus rate. As rate increases, the biphasic motor response becomes dominant resulting in the observed frequency doubling. From these data, we may conclude that the reorganizations observed in the brain dynamics are a reflection of the change in the relationship between the brain signals underlying the generation of movement and those arising from auditory processing.

Frequency dependent decreases in the amplitude of the average auditory evoked response have been reported during sensorimotor synchronization (Mayville et al. 2001) and auditory perception in the absence of movement (Carver et al. 2002). The relationship between response amplitude and tone interval is well established for long intervals on the order of 1 or more seconds (Hari et al. 1982; Lu et al. 1992; Sams et al. 1993). Carver et al. (2002) investigated the auditory response at frequencies ranging between 0.8 and 8Hz (ISI = 1250–125 ms), a relevant parameter regime for sensorimotor coordination and rhythm perception and processing in general (Fraisse 1982). These authors reported both a cessation of the N1m and a transition from a transient to a steady state response at stimulus rates of approximately 2.0 Hz, a rate of demonstrable perceptual significance (Fraisse 1982; Drake and Botte 1993) as well as the typical rate at which transitions from syncopation to synchronization occur (Kelso et al. 1990). Such findings provoke questions as to whether behavioral transitions might be a consequence of a qualitative change in auditory processing (and perception) associated with the onset of a steady state response. Controversy exists as to whether steady state responses reflect summation of overlapping long latency responses, implying a shared functional anatomy, or whether responses occurring on these two timescales are represented by separate neural networks (Pantev et al. 1996; Gutschalk et al. 1999). Functional MRI and PET studies may offer some insight into this controversy (Pastor et al. 2002). In either case, dynamic changes in auditory processing could ultimately alter the nature of the auditory to motor coupling in a way that disrupts the syncopated pattern in favor of the synchronized one. The intriguing relationship between stimulus processing and behavioral dynamics may represent a potential line of inquiry that has received little attention thus far (e.g. Lagarde and Kelso 2006).

Finally, to better understand the rate-dependence of cortical motor-related processes, Mayville et al. (2005) investigated self-paced movements using a continuation paradigm. Twenty one different metronome pacing rates were included spanning the range of .5Hz to 2.5 Hz in steps of 0.1 Hz. After 20 pacing tones, the stimulus was turned off and participants continued moving at the designated rate. They show that the time course of the classic motor-evoked field (MEF) is tightly coupled to the behavioral response (as had previously been found by Kelso et al. 1998) and does not depend on rate *per se*. However, a second, weaker motor-related field, consistent with a readiness field, was observed to strongly decrease in amplitude at rates above 1Hz. Interestingly, this is exactly the rate at which subjects switch from a reactive mode of coordination, in which responses follow the stimulus and an anticipatory or predictive

pattern in which responses begin before the metronome beat (Engstrom et al. 1996). The topology of this readiness field and its gradual accumulation of field amplitude prior to movement is consistent with a bilateral SMA source (Cheyne and Weinberg 1989; Erdler et al. 2001). Its decrease (by more than half when the rate exceeds 1.0Hz) may signify changes in the degree of motor processing necessary to move rhythmically at faster rates.

6 Oscillatory Activity and Coherence

There is currently considerable interest in understanding cognitive and motor functions in terms of large scale neural integration mediated by synchronous activity across different frequency bands (e.g., Babiloni et al. 2005; Brovelli et al. 2004; Chen et al., 2003; Gross et al. 2002; Gross et al. 2005; Varela et al. 2001). This interest reflects the hypothesis that synchronous oscillations may support the coordination of activity across distributed specialized cortical and subcortical regions and thereby form the foundation for complex neural representations and cognition (Basar, 2005; Bressler and Kelso 2001; Crick and Koch 2003; Singer 1994; 2001). Dynamic changes in the synchronous activity of the brain are reflected in large scale EEG and MEG recordings of human neural function (Nunez 1995) and are typically characterized as relative decreases or increases in power within specific frequency bands (Neuper and Pfurtscheller 2001). Decreases in band power are thought to reflect event-related desynchronization (ERD) within localized neuronal ensembles indicative of a change from an idling to a processing state (Lopes da Silva and Pfurtscheller 1999; Pfurtscheller and Andrew 1999). Conversely, event related synchronization (ERS) is thought to indicate an inhibitory or deactivated state in which the localized neural area is not involved in cortical processing (Lopes da Silva and Pfurtscheller 1999; Neuper and Pfurtscheller 2001). Regional changes in the power of oscillatory brain activity is complemented by the analysis of coherence in oscillations between distributed brain areas; a measure used to characterize the degree of functional coupling between brain regions (Singer 1993;1994; Gerloff et al. 1998). In this section we will draw from research that investigates, both directly and indirectly, the relationship of oscillatory power and coherence in the brain to behavioral coordination and stability.

6.1 Desynchronization, Coherence and the Stability of Coordination

A growing number of studies have investigated the spatiotemporal pattern of movement-related oscillatory power and coherence associated with unimanual sensorimotor synchronization (e.g. Pollok et al. 2005). Sensorimotor synchronization between finger movements of the dominant hand and external sensory stimuli is accompanied by a decrease in oscillatory power in both

alpha and beta bands over bilateral sensorimotor/premotor and medial premotor cortical areas (Gerloff et al. 1998; Manganotti et al. 1998; Deiber et al. 2001; Toma et al. 2002). Localized decreases in power have been reported in conjunction with an increase in interregional coherence, particularly between sensors approximately overlying contralateral sensorimotor cortex and medial premotor regions such as supplementary motor area (Gerloff et al. 1998; Manganotti et al. 1998; Toma et al. 2002). Such findings are thought to reflect a task-related increase in functional coupling between neural areas and underscore the importance of integration within a motor-premotor circuit for coordination.

Whereas frontal and parietal areas play an accepted role in the control of action (Rizzolatti et al. 1998), the specific relationship between activity in these regions and measures of behavioral dynamics such as stability and pattern switching is less established. However, there is growing evidence that such a relationship may exist. For example, moving from simple to more complex sequences of finger movement results in increased coherence in alpha and beta bands between contralateral sensorimotor and medial premotor sites (Manganotti et al. 1998). This increased coherence is accompanied by a decrease in low (8–10 Hz) (Manganotti et al. 1998) and high (10–12 Hz) (Hummel et al. 2003) alpha power over bilateral sensorimotor and parietal regions. Additional evidence suggests that the level of desynchronization in alpha and beta bands during coordination is related to the stability of the pattern (Chen et al. 2003; Jantzen et al. 2001; Mayville et al. 2001). Using MEG, Mayville and colleagues (2001) investigated differences in power in several frequency bands while participants produced syncopated and synchronized coordination at rates that increased systematically from 1.0 to 2.75 Hz. Syncopation was associated with significantly less power in the low (15–20 Hz) and high (20–30 Hz) beta bands over contralateral and central sensorimotor sensor sites. Greater desynchronization within these bands suggests greater neural engagement and higher processing demands during the less stable syncopation task. Taken together these findings lay the foundation for an hypothesis concerning the relationship between coordinative stability and neural function: less stable patterns of movement may be associated with both increased engagement in localized brain regions (indicated by increased event-related desynchronization) and concomitant increases in the functional coupling between motor and premotor areas (suggested by the increase in coherence in the EEG and MEG).

Studies that investigate how movement-related changes in power and coherence are influenced by practice further support the connection between brain activity and the stability of behavior. Behavioral studies already show that the intrinsic stability of a coordination pattern can be altered through practice. Learning increases both the stability of a previously unstable pattern (Zanone and Kelso 1992; Zanone and Kelso 1997) and the availability of attentional resources (Monno et al. 2002; Temprado et al. 2002; Temprado and Laurent 2004; Zanone et al. 2001). In an MEG experiment, Jantzen, Steinberg

and Kelso (2001) trained participants to syncopate at their individual critical frequencies (frequency at which transitions from syncopation to synchronization typically occur). Initially, syncopation resulted in a significant reduction in alpha and beta power at sensors overlying bilateral sensorimotor and premotor areas when compared to synchronization. Training successfully increased the stability of syncopation thereby shifting the critical frequency and delaying transitions to synchronization. Behavioral improvement was accompanied by a reduction or elimination in power differences between syncopation and synchronization suggesting that alpha and beta power may reflect the stability of the performed pattern. Interestingly this reduction was observed not only at the practiced rate but also at lower rates suggesting that stability improvements generalize to all movement rates.

Learning–dependent modulation of task related coherence has also been reported in subjects who practiced a task in which two independent unimanual finger sequences had to be combined in order to form a new bimanual antiphase coordination pattern (Andres et al. 1999; Gerloff and Andres 2002). Early in learning significant coherence in alpha (8–12.9 Hz) and beta (13–30.9 Hz) bands was observed between right and left sensorimotor areas and between these regions and mesial premotor cortex (SMA) (Andres et al. 1999). Such interregional functional coupling was greatly reduced at a later learning stage when performance was also found to be significantly more stable.

Although provocative, research findings suggesting that oscillatory power and coherence in motor and premotor areas are influenced by stability and changes in stability as a result of learning are still quite preliminary. So far, they allow one to infer a relationship between changes in neural activity and changes in stability in the case of learning a single coordination pattern only. Learning, however, has been shown to alter the dynamics of the entire system by either increasing or decreasing the accuracy and stability of coordination across a broad repertoire of patterns beyond those explicitly practiced (Kelso 1995; Schöner et al. 1992; Zanone and Kelso 1991; Zanone and Kelso 1992). Moreover, the nature of the change due to learning arises from the cooperative or competitive interplay between the pattern to be learned and pre-existing tendencies or "intrinsic dynamics" (Zanone and Kelso 1992). Understanding the neural basis of this interplay may provide greater insight into how stability and changes in stability are represented by the ongoing activity within and between a network of brain areas that minimally includes M1 and SMA (a broader network will be described in sect. 7).

6.2 Rate Dependence of Oscillatory Power and Coupling

Important insight into how localized oscillatory power and functional coupling is moderated by different rates of coordination is provided in a recent study (Toma et al. 2002). The brain dynamics was assessed on a rapid time scale to allow for a quantification of changes in power and coherence over the course of

a single movement cycle. Such work complements the majority of research in this area in which power and coherence measures are integrated over a longer time scale that typically includes multiple movement cycles. Thumb movements were synchronized with an auditory tone presented at 0.5, 0.75, 1, 2, 3 and 4 Hz. At slow movement rates (1 Hz and below), a transient pattern of power changes in both alpha and beta bands was observed over bilateral primary sensorimotor cortex and SMA. Desynchronization was evident just prior to and following the onset of EMG activity. This was followed by a relative increase in synchronized activity reflecting post movement rebound (Lopes da Silva and Pfurtscheller 1999). Cortical desynchronization was accompanied by a transient increase in coherence that quickly returned to or dropped below baseline levels prior to the start of the next movement cycle. This temporal progression of activity suggests that at low rates (where coordination is known to be multistable) successive coordinated movements are characterized by a phasic pattern of localized activation and deactivation accompanied by interregional coupling and decoupling. At rates of 2 Hz and above, a different temporal pattern of activity emerges: activity in alpha and beta bands remains desynchronized with respect to baseline for the entire movement cycle. Moreover, the functional coupling between motor and premotor regions remains elevated during the entire movement cycle.

The importance of the Toma et al (2002) study lies in uncovering the temporal evolution of neural interactions over the course of a single action-perception cycle and how such interactions change under increases in the control parameter of rate. The results show that during the synchronization task a qualitative change in the transient pattern of desynchronization and interareal coupling occurs at about the same frequency at which spontaneous switches between coordination states are observed ($\approx$ 2 Hz). The inclusion of additional, intrinsically less stable patterns such as syncopation in future work would allow for a better determination of the specific role that changes in motor-premotor coupling play in mediating stability and switching between patterns. The results of such studies may also offer an explanation for the intrinsic differences between coordination patterns in terms of altered coupling within specific neural circuits.

6.3 Differential Roles for Alpha and Beta Oscillations

Previous work has suggested a functionally distinct role for movement related alpha and beta desynchronization. For discrete self paced movements, beta or mu desynchronization (20 Hz) has been associated with activity of the primary motor cortex (i.e. precentral gyrus) and alpha desynchronization (10 Hz) with activity of somatosensory cortex (Salmelin et al. 1995). Definitive evidence determining whether alpha and beta desynchronization reflect dissociable processes during sensorimotor coordination is still lacking, however. For example, an increase in desynchronization during performance of the less stable syncopated pattern of coordination is observed in both alpha and beta bands

(Jantzen et al. 2001; Mayville et al. 2001). Moreover, the spatial distribution of this desynchronization is similar for both bands suggesting that alpha and beta desynchrony may reflect a single underlying process. Although Toma and colleagues (Toma et al. 2002) reported rate related changes in oscillatory power and coupling in both alpha and beta bands, their experimental effects were clearly more pronounced in the beta band. In contrast, Pollok et al. (2005) reported that coherence between distributed brain regions involved in auditory-motor synchronization was mediated primarily within the alpha range with relatively little beta contribution. However, because only synchronization was investigated in the latter study, the possibility that beta desynchronization increases when performing less stable patterns of coordination could not be assessed.

Recent MEG work by Chen, Ding and Kelso (2003) suggests different functional roles for alpha and beta bands. Subjects performed a combination of four conditions that allowed for a distinction between patterns of activity related to kinematic (flexion/extension) and relative timing (syncopation/synchronization) demands. Alpha activity was associated with the presence or absence of movement with similar desynchronization observed across all conditions. In contrast, beta activity distinguished *between* patterns of coordination: synchronization conditions revealed an increase in task related power in left sensorimotor and frontal regions whereas syncopation resulted in beta power decreases in contralateral sensorimotor areas. These findings suggest that alpha reflects neural processes associated with making finger flexion-extension movements while beta reflects more abstract or cognitive demands associated with the organization of different coordination patterns. Such an hypothesis is supported by a recent EEG study from Deiber and colleagues (2005). These authors employed a bimanual coordination task and assessed cortical activity following a precue that specified information concerning the coordination pattern ("abstract" feature), the fingers to move ("concrete" feature) or both. Conditions that allowed for pre-planning of the pattern demonstrated the greatest beta (15–25 Hz) desynchronization. Since such desynchronization was observed prior to movement onset, it may reflect processes related to the organization of the coordination pattern more so than those occurring during the movement itself.

Taken together, these results show that the neural activity manifest by a change in oscillatory (alpha and beta) power is related to the organization of coordinated action. Both bands reflect differences in stability between coordination patterns and track learning induced changes of intrinsically unstable patterns. Such neural activity may be related to the organization of abstract features of coordination at the level of the stability of the spatiotemporal pattern produced as opposed to the activation of the fingers *per se*. Although EEG lacks a high degree of spatial specificity, the spatial distribution of beta ERD implicates bilateral sensorimotor, premotor and supplementary motor areas in this process suggesting that these regions form a network critical to the organization of coordinated action. Although methods to extend and

improve spatial resolution and localization accuracy of EEG and MEG are under development (see sect. 8.0), the inherent limitations of these two measurement techniques restrict their effectiveness for investigating and identifying multiple distributed neural sources, particularly those in the midbrain and cerebellum. As a consequence, subsequent research has turned to functional imaging approaches such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) to identify and characterize the distributed neural networks supporting sensorimotor coordination. We turn to a consideration of these next.

7 Brain Networks Underlying Patterns of Coordination Revealed Using fMRI and PET

A first objective is to identify the basic cortical and subcortical networks that support sensorimotor coordination and to determine whether these networks differ for different behavioral patterns. Using fMRI, Mayville and colleagues (Mayville et al. 2002) sought to determine whether different brain networks underlie the two dominant coordination modes, synchronization and syncopation. During coordination conditions participants were required to either syncopate or synchronize with an auditory metronome presented at a constant rate of 1.25 Hz. At this slow rate both synchronization and syncopation can be performed accurately. Controls included a listening condition in which participants heard the metronome but did not move, and a motor only condition during which rhythmic self-paced movements were made in the absence of an auditory metronome. All experimental conditions were presented in a standard block design of alternating periods of task and rest during which no stimuli were presented and participants remained still.

A contrast between coordination modes revealed a large-scale brain network that was more active for the intrinsically less stable syncopated coordination pattern. Syncopation was associated with either increased or additional activity in bilateral dorsal premotor cortex, supplementary motor area, insula, bilateral cerebellum and bilateral basal ganglia and thalamus. No areas demonstrated greater activity during synchronization compared to syncopation. Individual subject results suggested that activity in primary sensorimotor cortex, SMA and ipsilateral cerebellum was common to both coordination modes since the majority of participants showed activity in these areas during both tasks. However, activity in the SMA and cerebellum was greater for syncopation. In contrast, activity in lateral premotor cortex, basal ganglia and contralateral cerebellum was observed in only a small subset of subjects during synchronization. The Mayville et al (2002) results confirm previous EEG and MEG findings that found coordination dependent modulation of activity over premotor areas (e.g. Fuchs et al., 1992, 2000b). In addition, they target a broader set of cortical and subcortical regions for the less stable coordination pattern.

Follow-up fMRI studies have replicated and extended the foregoing findings, consistently identifying syncopation dependent increases in blood oxygen (BOLD) in a network that includes SMA, dorsal premotor cortex, basal ganglia, insula and cerebellum (Jantzen et al. 2002; 2004; 2005; Oullier et al. 2005). This work has demonstrated that the BOLD differences identified by Mayville et al. (2002) are not strictly related to differences in sensory and motor demands of the two patterns, but instead may be related to higher-level cognitive processes associated with the formation of the required coordination pattern. In one of these studies, the importance of explicit motor output and the resulting sensory feedback in determining coordination dependent differences between synchronization and syncopation was assessed using an imagination paradigm (Oullier et al. 2005). The primary finding was that compared to synchronization, syncopation resulted in significantly greater BOLD signal amplitude in SMA, premotor cortex, thalamus and contralateral cerebellum regardless of whether coordination was actually performed or only imagined (i.e. in the absence of any movements). This finding may be taken as evidence that activity in these brain areas is related to processes critical for organizing the pattern of coordination regardless of whether it is actually performed or not. Activity in primary motor cortex and ipsilateral cerebellum, on the other hand, was much greater for actual performance conditions, as might be expected given their proposed role in purely sensory and motor aspects of coordination.

7.1 Parametric Approaches Connecting Rate, Stability and BOLD

The foregoing studies employed fMRI to investigate the brain networks underlying synchronized and syncopated patterns of coordination when performed at a single, slow rate. Results revealed a distributed network of areas that are more active in response to syncopation when compared to synchronization. But what is the role of this brain network? How can activity within this network be related to the pattern dynamics of behavioral coordination expressed through the collective variable (relative phase) and its stability? The beauty of the 'nonlinear paradigm' is that stability can be experimentally manipulated by parametrically controlling the rate of movement for different coordination patterns. Thus, increasing rate during synchronization has little or no effect on the stability of coordination whereas similar increases during syncopation result in a gradual decrease in stability, eventually leading to large scale switching between patterns. It would seem, therefore, that a critical link for tying together behavioral pattern formation and brain activity is to determine how activity within and between specific brain networks is altered by systematic parametric variation in coordination rate (and hence stability).

Rate

The majority of studies investigating the relationship between measures of neural function and changes in the rate of movement or stimulus presentation

have emphasized the dissociation between brain areas that are task dependent (i.e., demonstrate similar increases in activity regardless of rate) and areas that are rate dependent. There is widespread agreement for a positive linear relationship between rate and level of activity observed in primary visual (Fox and Raichle 1984), auditory (Binder et al. 1994; Price et al. 1992; Rees et al. 1997) and sensorimotor (Jancke et al. 1998; Jenkins et al. 1997; Kawashima et al. 1999; Rao et al. 1996; Sabatini et al. 1993; Sadato et al. 1997) regions. In this latter group of studies, activity in primary somatosensory and motor regions (i.e. pre and post central gyrus) has been shown to increase in a monotonic (and in most cases linear) way with systematic increases in the rate of sensorimotor synchronization. However, reports are mixed when considering rate dependence across broader sensory and motor networks that include premotor, insular and cerebellar regions. Inspection of the literature reveals that such discrepancies likely stem from experimental factors and may be attributed primarily to differences in the nature of the motor tasks employed and, to a lesser degree, differences in the range of rates investigated.

In studies employing relatively simple movement paradigms, rate dependent activity has been restricted to primary perceptual-motor areas and their dominant pathways. For instance, Jancke and colleagues (Jancke et al. 1998) measured BOLD signal changes in response to simple button presses synchronized with a visual stimulus presented at rates from 0.5 to 5.0 Hz in 0.5 Hz increments. Activity was positively correlated with movement rate in contralateral sensorimotor cortex and posterior parietal cortex, both of which are likely related to motor and sensory (dorsal visual stream) processing. In a study of self-paced finger tapping, with movements performed at $^1/_4$, $^1/_2$, 1, 2 and 4 times the subject's preferred rate, related CBF increases were confined to contralateral primary sensorimotor cortex (Kawashima et al. 1999). Although SMA, parietal,and cingulate cortex were all activated during the task, their level of activity did not scale with rate. These data tend to support the notion that activity in primary input-output regions is modulated by the basic motoric demands associated with increasing the number of movements per unit of time, whereas activity in other areas is associated with processes that are not directly altered by rate. Interestingly, in the Kawashima et al. (1999) work, the non-rate dependent areas demonstrated increased blood flow for all rates other than the preferred frequency, regardless of whether movements were faster or slower. Such a finding suggests that activity in such areas may be more responsive to movement features such as difficulty, stability or variability as opposed to rate *per se*.

In contrast to simple motor tasks, more elaborate movements or movements performed at fast rates (> 4.0 Hz) reveal a rate dependency expressed over a much broader network of functional areas (Jancke et al. 1998; Jenkins et al. 1997; Riecker et al. 2003; Sabatini et al. 1993). For instance, when finger to thumb opposition movements are performed in place of simple finger tapping, a rate dependency is observed in SMC as well as in SMA and bilateral cerebellum (Sabatini et al. 1993). Jenkins and coworkers (1997) employed a

coordination task in which participants were free to make joystick movements in any chosen direction. These multi-faceted movements (requiring coordination across multiple joints) were paced with an auditory signal presented at rates between 0.2 and 1.0 Hz. This paradigm was associated with a positive linear relationship between rate and cerebral blood flow (CBF) within SMC, SMA, cerebellum, premotor cortex, insula and primary auditory cortex.

At higher movement rates two different relationships are found depending on the functional area under question. Reicker et al., (2003) had participants perform synchronized tapping at rates from 2.0 Hz to 6.0 Hz, frequencies typically at or above the rates at which syncopation can be stably performed (Kelso et al. 1990) and that approach the biomechanical limit for producing rhythmic movements. Linear rate related increases in BOLD were observed in contralateral SMC, SMA proper, pre-SMA and left thalamus. A step-wise increase from low (< 4.0 Hz) to high (> 4.0Hz) rates was also observed in two areas of the ipsilateral cerebellum. This increase occurred at the same frequency at which activity in SMC appeared to reach an asymptote suggesting a qualitative change in processing and production at these high movement rates.

Taken together these data support two important conclusions concerning the role of movement rate in modulating neural activity. First, the combined evidence from a large number of studies strongly supports the conclusion that activity in primary sensorimotor cortex is modulated by the speed at which movements are performed, at least up to rates below those at which biomechanical influences are felt. Second activity in non-primary motor areas including lateral premotor cortex, SMA, cerebellum and the insula are not influenced by movement rate *per se*, but likely reflect changes in more abstract task features such as difficulty or complexity that arise from the combined influence of task demands and movement rate.

Behavioral Complexity

More recently there has been an explicit attempt to understand the specific link between the complexity of motor behavior and levels of brain activity. Such studies serve to further dissociate brain regions that respond differentially to parameters such as rate, force, complexity and so forth (Wexler et al. 1997). It is necessary to clarify here that although the term "complexity" has been used to imply a specific conceptual, theoretical or mathematical approach to understanding human action in terms of emergence, pattern formation, self organization and the like (e.g. Coveney and Highfield 1995; Kelso 1995) and also in terms of neural complexity (e.g. Sporns, 2004; Tononi, et al., 1998), in the literature cited in this section, "complexity" is employed in a more colloquial way to refer to the categorization of movements according to the difficulty people have producing them. Consequently, the specific features manipulated in altering the complexity of a movement vary across the literature according to how complexity is conceived. Complexity has often been

explored by employing tasks that involve the spatiotemporal ordering of specific sequences of finger movements of the same hand (Harrington et al. 2000) or of movements of the two hands (Tracy et al. 2001; Meyer-Lindenberg et al. 2002; Ullen and Bengtsson 2003; Debaere et al. 2004). In the unimanual case, manipulation of complexity has been achieved by increasing the number of movements performed (Catalan et al. 1998), by altering the relative timing between movements (Dhamala et al. 2002; Lewis et al. 2004) or by adjusting the ordering between fingers with less sequential orderings being considered more complex (Haaland et al. 2000). Varying complexity on a number of arbitrary dimensions has resulted in a broad interpretation of the functional role of the brain areas involved.

Initial studies explored the role of predictability in accounting for activity in premotor networks. Predictability of a motor sequence has been associated with increased activity in a network of brain regions (VanOostende et al. 1997; Dassonville et al. 1998) that overlap with those sensitive to rate (Sabatini et al. 1993; Jenkins et al. 1997; Jancke et al. 1998; Riecker et al. 2003) and those distinguishing between coordination patterns (Mayville et al. 2002). When compared to predictable cued finger sequences such as repetitively moving fingers in a fixed order, random and therefore unpredictable movements were associated with increased BOLD activity in premotor cortex, SMA, pre-SMA and parietal cortex (VanOostende et al. 1997; Dassonville et al. 1998). These empirical results have been interpreted in terms of a link between activity in frontal and parietal regions and cognitive operations associated with decreased predictability (i.e. greater demands on planning, decision making and memory). However, in terms of the ordering between fingers, the predictable condition always required a simple sequential pattern whereas the unpredictable condition required transitions between non-adjacent fingers.

An alternative explanation is that activity in premotor and parietal regions was influenced by changing the spatiotemporal pattern to be produced in the two conditions. Ordered movements that produce simple patterns where fingers are moved in increasing or decreasing order (e.g. 2345) are less complex than patterns requiring coordination between non-adjacent fingers (3425). This alternative view has been supported by subsequent EEG work that independently varied predictability of a sequence and sequence order (Hummel et al. 2003). These authors showed that task related power decreases, a measure of cortical engagement thought to reflect local neural resources (Lopes da Silva and Pfurtscheller 1999), were modulated by the complexity of the spatiotemporal pattern produced, but were similar across conditions that varied in predictability and memory load.

In keeping with the foregoing research is a study in which movement patterns were reformulated in terms of muscular synergies in which synergistic movements occur when fingers flex together in a synchronized (in-phase) fashion and non-synergistic movements involve alternating flexion of digits in a syncopated (anti-phase) fashion (Ehrsson et al. 2002). Nonsynergistic movements were associated with greater activity in SMA, cingulate motor

area (CMA), bilateral dorsal premotor cortex, intraparietal sulcus and lateral cerebellum. This pattern of neural activity is remarkably similar to the one resulting from comparisons between synchronized and syncopated coordination (Mayville et al. 2002) where behavior is defined in terms of relative phase and the pattern of muscle recruitment remains the same throughout. It is no secret that one of the hallmarks of synergies is 'invariant' or stable timing relations among component elements across parametric change (Kelso 1986; 1997).

The relative importance of the coordination between fingers as opposed to predictability, rate or memory has been further supported by a group of fMRI studies in which motor sequences were over-learned prior to the onset of scanning. Catalan et al. (1998) increased complexity by increasing the number of fingers used to perform a sequence and by altering the movement order and length of the sequence to be performed. Their PET results showed that increasing complexity was associated with increases in CBF in SMA, premotor cortex, cerebellum, bilateral parietal cortex and precuneus. A similar approach in which finger sequence length and complexity were defined with respect to the order of fingers moved was employed in PET (Boecker et al. 1998) and fMRI (Haslinger et al. 2002) studies. All movements were well practiced and movement rate, force and the overall number of movements performed were maintained across conditions. The fMRI results revealed clusters of complexity related BOLD activity in bilateral parietal cortex, left premotor cortex and the dentate nucleus in the right cerebellar hemisphere (Haslinger et al. 2002), further implicating the cerebellum in this complexity related network. The PET results suggest a role for central midline structures, showing a relationship between increased complexity and greater activity in the SMA, pre-SMA, and bilateral basal ganglia (Boecker et al. 1998). Taken together, this body of work supports the hypothesis that frontal motor networks and possibly cerebellum play an important role in mediating coordination in functional tasks of increasing complexity.

A further distinction has been offered by Harrington et al. (2000) who sought to directly differentiate the role of physical (surface) features of the sequence task and more cognitive or abstract aspects. Of relevance to the current review is that activity in dorsal premotor cortex, middle frontal gyrus, supra marginal gyrus and left insula was associated with the number of transitions between fingers, independent of the number of fingers used to generate the sequence (3 vs. 2). Such a finding suggests that these brain areas mediate processes central for guiding the coordination between fingers, that is, for generating the required spatiotemporal pattern. In contrast, a network comprised of superior parietal lobe and contralateral cerebellum was found to be active exclusively in response to changes in the number of fingers employed. However, two curious findings emerged from this important study. First, complexity did not modulate activity in SMA, a region commonly reported to be responsive to changes in complexity (Harrington et al. 2000). Second, in contrast to most of the literature, a negative relationship was found between BOLD amplitude and complexity in a majority of the regions identified. The Harrington et al.

work provides important methodological and conceptual advances for understanding complexity in terms of specific motor demands and task dimensions, specifically the spatial and temporal relation between individual movements. However, their somewhat discrepant results underscore the need for further investigation of this relationship with particular focus on how specific task features are represented at the neural level.

Stability

Research linking parameters such as rate and concepts such as complexity to changes in activity within (and across) functional networks of the brain is provocative because it suggests that brain areas supporting basic sensory and motor aspects of coordination (i.e. primary sensory and motor areas) might be distinguished from those involved in processing more abstract demands (i.e. SMA, premotor cortex and cerebellum). Such research has important implications for understanding coordinated behavior and leads us to an exciting hypothesis regarding the neural circuitry underlying syncopation and synchronization, namely that the basic neural circuitry for processing sensory and motor elements of a task may be separable from the neural circuitry responsible for mediating the coordination between component elements. The former network may be expected to reflect changes in rate and not be sensitive to intrinsic differences in stability between coordination patterns. In contrast, the network of areas identified by Mayville et al. (2002) and others may underlie the stability of coordinated behavior regardless of the particular sensory and motor elements involved.

Within the Coordination Dynamics framework, relative phase (with its multi- and metastable properties) provides a low dimensional description of the collective spatiotemporal ordering between coordinating elements. This has led some to hypothesize that changes in stability measured by the parameter of relative phase may be reflected in networks that differentiate between in-phase and antiphase modes of coordination (Meyer-Lindenberg et al. 2002). This hypothesis leads to the prediction that activity in networks concerned with the relative phase between components should track the variability (inverse of stability) of this parameter. Since decreases in stability with increasing movement rate are observed during antiphase and not in-phase coordination, increases in neural activity within this network should be observed only for the former pattern. However, areas active in response to changes in movement rate should respond similarly regardless of pattern.

To address these predictions recent studies have investigated the relationship between large scale measures of brain function and the stability of bimanual coordination using PET, TMS (Meyer-Lindenberg et al. 2002) and fMRI (Debaere et al. 2004). In the former work cerebral blood flow was measured while performing in-phase and anti-phase movements at four movement rates (1.0, 1.5, 1.7, 2.0 Hz) below the critical frequency (Meyer-Lindenberg et al. 2002). Moving at increasingly higher rates resulted in a significant reduction

in stability only for antiphase movements, a finding compatible with well-known studies of bimanual coordination and theoretical predictions thereof. Increases in measures of cerebral blood flow concomitant with increasing instability were observed in bilateral dorsal premotor cortex, SMA and cingulate as well as the left frontal operculum, ipsilateral cerebellum and left supramarginal gyrus. Activity in contralateral sensorimotor cortex, on the other hand, demonstrated a linear increase in CBF for both patterns indicating sensitivity to changes in rate regardless of pattern stability.

Comparable changes in BOLD were reported when movements were made either in-phase, anti-phase or in a 90 degree pattern (Debaere et al. 2004). For the latter performance was measured in terms of relative phase error. Error may not provide an optimal index of stability since, when averaged over a relatively long period of time, even unstable coordination patterns can appear to have little or no phase error. Nonetheless, a monotonic decrease in performance with increasing cycling frequency was observed for the antiphase and 90 degree patterns but not for the in-phase pattern. As expected, both dorsal premotor cortex and cerebellum demonstrated an interaction between movement rate and coordination pattern, showing a BOLD increase for the out of phase and 90 degree patterns and not for the in-phase pattern.

We recently performed a similar parametric study to investigate the mapping between BOLD amplitude, movement frequency and stability of unimanual sensorimotor coordination patterns (Jantzen et al., submitted). Participants coordinated right hand finger flexion with an auditory metronome in either a synchronized or syncopated pattern at five different movement rates (0.75, 1.0, 1.25, 1.50, 1.75 Hz), all below the critical frequency at which spontaneous transitions from syncopation to synchronization are known to occur. Changes in stability-related brain regions induced by parametric manipulation of coordination frequency should be mirrored as systematic increases in BOLD signal intensity during syncopation but not during synchronization since stability remains relatively unaffected for the latter. Theoretical considerations (see Sect. 4) further predict that regions sensitive to systematic increases in rate or frequency should exhibit monotonic increases in signal intensity independent of the coordination pattern performed.

The results provide neurophysiological evidence for the existence of two separate but interacting brain networks underlying basic patterns of perceptual-motor coordination. BOLD amplitude in contralateral SMC and bilateral primary auditory areas increased linearly with increases in movement rate for both syncopated and synchronized coordination (Fig. 3). This pattern of activity is consistent with a modality dependent processing role focused primarily on the elementary sensory and motor demands of the task. In contrast, a second network consisting of pre-SMA, bilateral premotor, ipsilateral cerebellum and right insula, exhibited activity that mirrored behavioral stability. In keeping with our theoretical framework, BOLD amplitude in this network of brain areas increased linearly with decreases in the stability of the syncopate

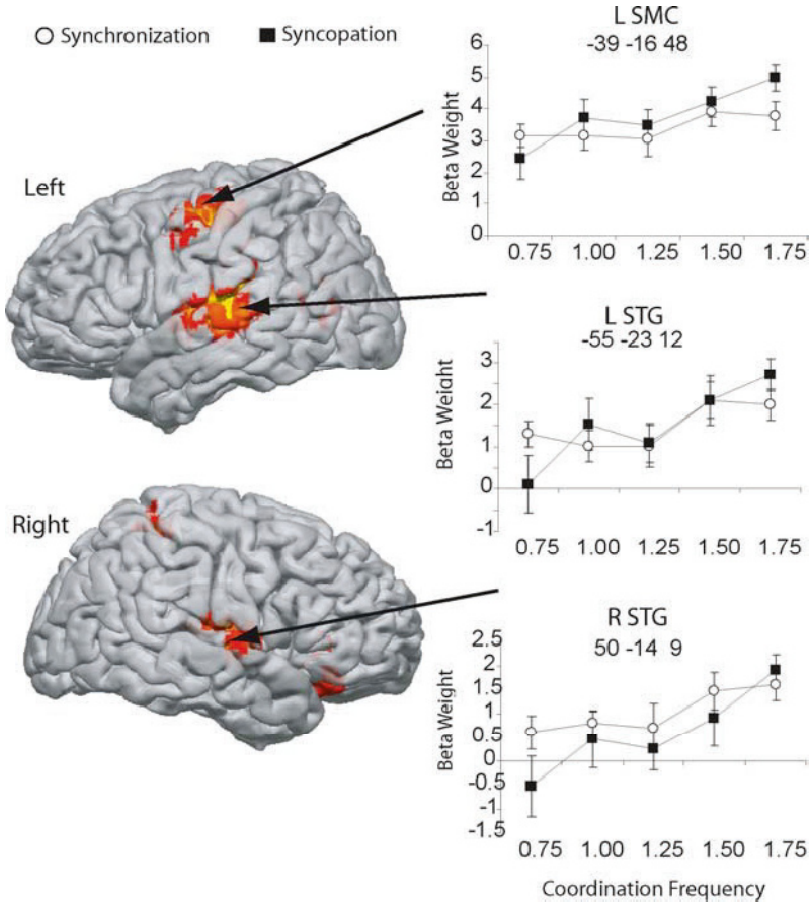


Fig. 3. Cortical regions exhibiting a main effect of movement rate/coordination frequency are plotted in color on a template brain rendered in 3 dimensions and viewed from the left (top) and right (bottom) lateral views. The between subject mean of the regression coefficients (beta weight) for each experimental condition is plotted as a function of the required coordination frequency. BOLD amplitude patterns were very similar for both synchronization (open circles) and syncopation (closed squares) suggesting that these brain areas are related to basic sensory and motor demands regardless of mode of coordination.The brain areas identified as well as the Tailarach coordinates are presented in each plot. Error bars are set to ± standard error

pattern (Fig. 4). No such increases were observed during synchronization since this pattern was observed to remain stable across movement rate.

The parametric approach of coordination dynamics thus allows for the differentiation between brain networks related to a) dynamic features of coordination that represent the global organization of the behavioral pattern, in particular its stability and b) those related to more elementary modality

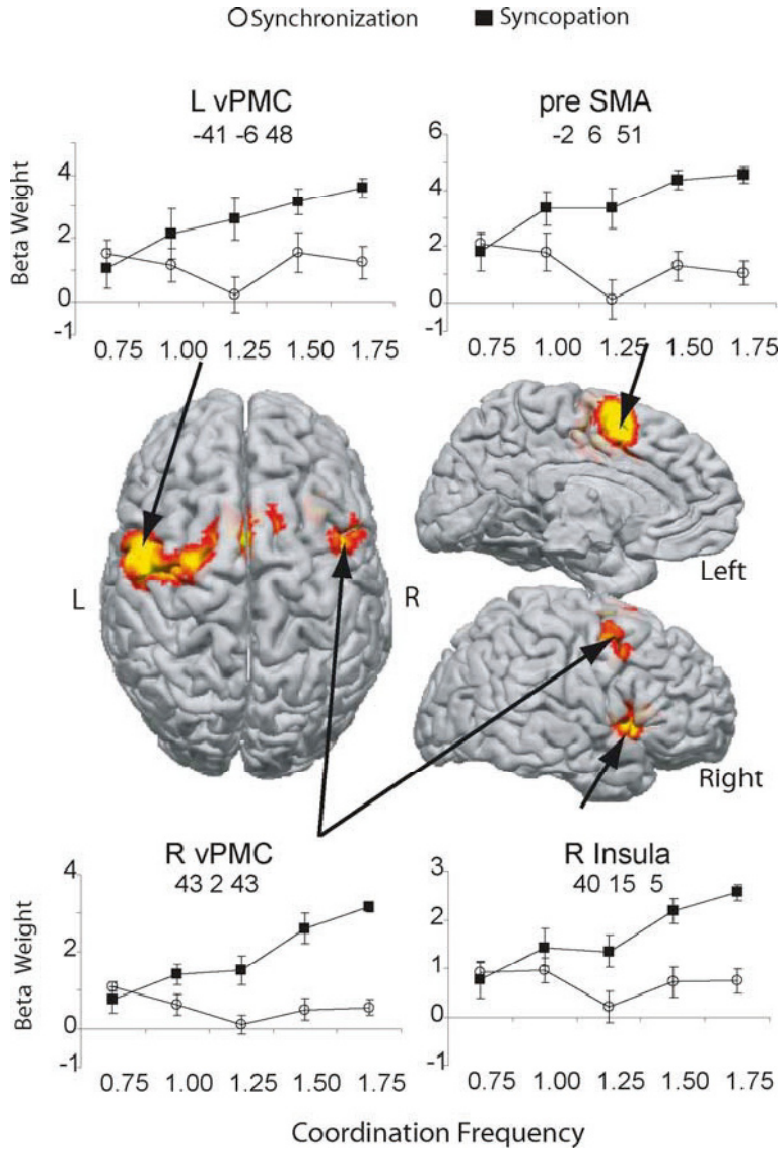


Fig. 4. Regions exhibiting an interaction between movement rate/coordination frequency and sensorimotor pattern are plotted in color on a template brain rendered in 3 dimensions. Views are of the dorsal surface of the brain viewed from the top, and the left medial and right lateral surface of the brain both viewed sagitally. Plotting conventions are as in Fig. 3. Here a clear dissociation between synchronization and syncopation is observed in medial and lateral premotor areas as well as in the right insula. The pattern of BOLD activity closely follows the pattern of behavioral variability, a measure of stability, by increasing monotonically with increasing rate during syncopation but not during synchronization. Thus, these areas respond in relation to differences in pattern stability as opposed to movement rate per se. vPMC: ventral premotor cortex., preSMA pre supplementary motor area

dependent sensory and motor features. Studies using transcranial magnetic stimulation (TMS) have provided further support for a link between the stability of coordination patterns and activity in cerebellar-frontal networks (Meyer-Lindenberg et al. 2002; Serrien et al. 2002; Steyvers et al. 2003). TMS was used to transiently disrupt SMA and lateral premotor cortex during performance of bimanual coordination. Applying TMS to these regions during antiphase coordination resulted in increased phasing errors between the hands (Meyer-Lindenberg et al. 2002; Serrien et al. 2002; Steyvers et al. 2003) and induced behavioral transitions from anti-phase to in-phase coordination (Meyer-Lindenberg et al. 2002). No like alterations in behavior were observed when TMS was applied during in-phase coordination. Moreover, the magnitude of the TMS perturbation needed to provoke a behavioral transition decreased as behavioral instability increased (Meyer-Lindenberg et al. 2002). Similar TMS disturbances applied to primary motor cortex did not affect the pattern of coordination indicating that the results were not due to a simple perturbation of the efferent signal (Meyer-Lindenberg et al. 2002). Overall the foregoing results suggest the following hypothesis: Activity across a network that minimally includes SMA, lateral premotor cortex and cerebellum is linked to the degree of behavioral stability. Depending on how close the brain is to an instability, disruption of this network may result in the destabilization and eventual dismantling of a less stable pattern in favor of a more stable one.

7.2 The Generic Nature of Coordination Dynamics and Stability Dependent Networks

The research reviewed here supports the notion that key collective variables or order parameters that capture coordination at the behavioral level are also critical for representing coordinated action at the level of the brain. New findings illustrate how such pattern formation and change is mediated by the integrated activity of cortical and subcortical networks. An important feature of relative phase dynamics is that it provides an effector- and modality–independent description of the emergence of coordination patterns. This begs the question: Is there an effector or modality independent representation of this quantity at the level of the brain? While the present review has been focused on sensorimotor coordination, it appears that highly similar brain networks support coordination whether coordinating with oneself or with the external environment. The existence of such stability dependent networks leads inevitably to the proposal that common neural substrates may support behavioral coordination regardless of the coordinating elements themselves.

Considered as a whole, the imaging work underscores the importance of the SMA and premotor cortices in mediating coordinative stability independent of the specific elements being coordinated. There is growing evidence for stability dependent networks regardless of whether the patterns of behavior refer to the fingers of a single hand (Ehrsson et al. 2002; Nair et al. 2003), between oneself and the environment (Jantzen et al. 2002; Mayville et al. 2002;

Jantzen et al. 2004; Jantzen et al. 2005; Oullier et al. 2005), between homologous (Sadato et al. 1997; Immisch et al. 2001; Meyer-Lindenberg et al. 2002; Ullen and Bengtsson 2003; Debaere et al. 2004) and non-homologous limbs (Debaere et al. 2003) . Stability dependent networks are engaged even when coordination is imagined (Nair et al. 2003; Oullier et al. 2005). Such evidence speaks compellingly for the tenet that, like behavior itself, the coordination dynamics of the human brain can be captured by informationally relevant quantities such as the relative phase among interacting coordinating elements (Kelso 1994; Deiber et al. 2005). The specific neural representation of the coordinating elements or agents will vary in task and modality dependent ways. In the examples described here, activity of primary sensory and motor areas is related to the elements being coordinated independent of the pattern of coordination. However, an emerging insight is that other brain regions form a network that can be differentiated from modality specific areas by virtue of their relationship to the underlying coordination dynamics.

8 Summary and Future Directions

In the last 20 years or so, sensorimotor coordination has proven itself a cardinal example of self-organizing, pattern forming processes and, by virtue of that fact, has provided an ideal experimental platform for studying the dynamics of brain and behavior. In this chapter we have reviewed literature relevant to forming a conceptual and theoretical link between dynamic aspects of behavioral coordination and spatio- temporal properties of brain function. Taking the theoretical perspective that coordination in complex biological systems is fundamentally a result of self-organizing dynamics, we have discussed research that describes how interactions among coordinating elements is mediated at the level of the brain and how such brain activity may underlie specific dynamic features of behavioral coordination such as loss of stability and switching between patterns. Evoked potential studies of auditory-motor coordination have highlighted the possible role of the dynamic interplay between auditory and motor related activity in the brain. Transitions between behavioral patterns occur at approximately the same frequency as the onset of auditory steady state evoked responses. This intriguing finding has received little subsequent inquiry and additional investigations are required to firmly establish the role of primary perceptual areas in mediating rate-dependent coordination stability and switching.

A relationship between the stability of the coordination pattern and activity within a fronto-cerebellar network has also been established based on converging evidence from multiple imaging studies. Activity in a network that includes SMA, lateral premotor cortex, insula and cerebellum appears to be related to the stability of the relative phase dynamics, suggesting a role for this network in organizing and monitoring the global relationship between coordinating elements. This network (a) is more active for less stable coordination

patterns; (b) reflects changes in stability induced through learning; and (c) reflects changes in stability induced through manipulation of control parameters. This stability related network has been distinguished from primary sensory and motor processing areas that respond to changes in the rate of coordination but do not appear to be sensitive to alterations of pattern stability. Although detailed information about the interaction between areas that form these networks is still lacking, initial work suggests that less stable patterns are associated with increased coherence between SMA and M1.

A schematic representation of the rate and stability related networks and the putative connections between them is provided in Fig. 5. The critical next steps in understanding the relationship between cortical and subcortical networks underlying dynamic properties of coordination will focus on the pattern of neural coupling between individual brain areas and how that pattern relates to coordination. What are the dynamics of the two networks and how do they interact? It is now widely accepted that cognition and behavior emerge from the dynamic interplay between the dual tendencies of segregation, in the form of localized activity within functionally specific neural regions,

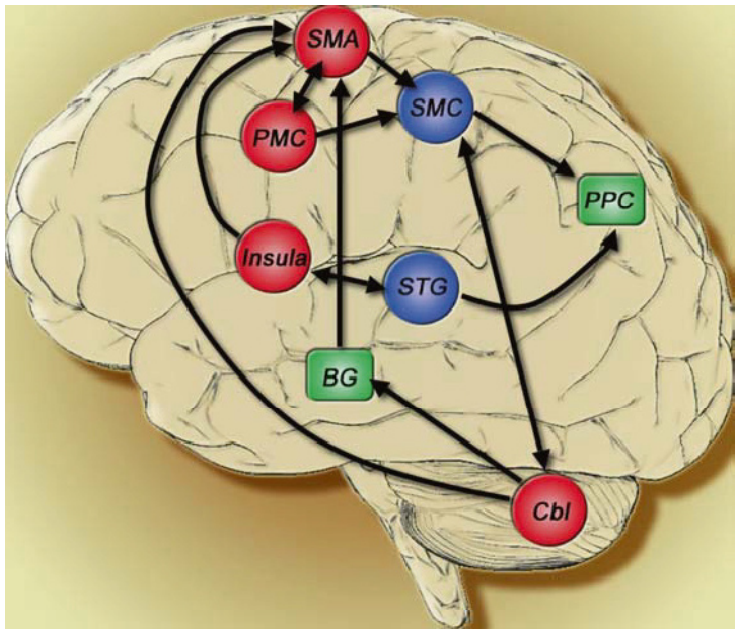


Fig. 5. A schematic representation of brain circuits supporting sensorimotor coordination and the primary functional connections between them (*arrows*). Blue and red areas demonstrate rate and stability dependent activity respectively. Green areas have been inconsistently observed in the literature and therefore their importance for rhythmic coordination is less established. SMA: supplementary motor area, SMC: sensorimotor cortex, PMC: premotor cortex, PPC: posterior parietal cortex, STG: superior temporal gyrus, Cbl: Cerebellum

and integration as realized by interregional communication between regions (see Bressler & Kelso 2001; Bressler & McIntosh 2007; Kelso 1995; Kelso & Engstrom 2006; Sporns & Tononi 2007). While the work represented in this chapter has made significant strides in identifying specific brain regions and patterns of brain activity corresponding to patterns of human behavior, much more research is needed to understand the networks themselves in terms of detailed structure, connectivity and dynamics — both in normal and clinical populations.

Advances in understanding neural dynamics and how this dynamics maps onto behavior will be aided by new analysis techniques that provide enhanced temporal and spatial resolution (Gross et al. 2001; Jirsa et al. 2002; Astolfi et al. 2005; Babiloni et al. 2005) and allow for a characterization of the directed influence between interconnected brain regions (Beggs, Clukas & Chen 2007; Breakspear & Jirsa 2007; Horwitz & Husain 2007; Stephan & Friston 2007). Spatial filtering techniques such as the variety of beamformer methods (Darvas & Leahy 2007; Fuchs 2007) provide a means to move toward brain-based measures of neural activity that possess both high spatial (mm) and temporal (ms) resolution. The promise of such approaches for combining results from multiple imaging modalities is demonstrated by the spatial overlap between desynchronization in beta band oscillations localized on the surface of the cortex and BOLD based imaging results generated during simple motor (Taniguchi et al. 2000), as well as language and motion detection tasks (Singh et al. 2002). Initial investigations have begun to probe the network properties supporting simple internally and externally paced movements (Gross et al. 2002; Tass et al. 2003; Babiloni et al. 2005; Pollok et al. 2005). Also, some advances have been made toward understanding how the coupling between a restricted network of brain areas changes as a function of different coordination patterns (Gross et al. 2005).

Questions critical to understanding the relationship between behavior and the spatiotemporal dynamics of the brain that are now open to new analytical approaches include:

- What is the nature of the functional interaction between brain areas forming hypothesized stability and modality dependent networks? How are neural interactions facilitated within and between these networks (see Fig. 5)?
- How are interactions within and between networks supporting different behavioral patterns altered by systematic changes in coordination pattern, rate and stability? Are losses in stability accompanied by a decrease/increase of coupling between brain areas forming these two networks? Alternatively, is coupling between brain areas altered by learning, intention or attention induced increases in stability of coordination?
- Can different functional networks be defined based on the oscillation frequencies of participating areas (e.g. alpha vs. beta vs. gamma), and if so how do these networks interact in space and time?

- What happens to established brain networks around critical points where stability is lost and new patterns adopted? Are the neural processes that underlie switches from syncopation to synchronization generalizable across different instances of behavioral pattern switching and decision-making?
- Does the brain contain generic networks for supporting dynamic properties of behavior, such as stability? Can a similar network be identified across multiple forms of coordination (e.g. multisensory, interpersonal, interlimb) independent of the coordinating elements (visual, auditory, somatic, etc) involved? Although there is some evidence that such is the case, this question has never been addressed exhaustively across experimental tasks particularly at the level of large scale brain networks.
- How is behavioral coordination achieved under conditions in which key connections or nodes of the network are damaged or disrupted as in traumatic brain injury, Parkinson's Disease, stroke, cortical or cerebellar lesions, etc.

Clearly the foregoing list of questions is far from inclusive, representing but a small sample of how recent advances in analytical approaches to neuroimaging data may be used to explore coordination. The questions posed are meant to provoke further research into establishing a direct connection between the theoretical and mathematical formulations of behavioral dynamics and measures of brain function on the one hand, and to understand how the rich dynamics of human behavior arise from the large-scale interaction among distributed brain areas that share a degree of specialization on the other.

Acknowledgments

Our thanks go to Julien Lagarde, Olivier Oullier and Paul Ferrari for comments on earlier versions of this manuscript. Much of our own work presented in this paper was supported by NIMH grants 42900 and 01386. The preparation of this manuscript was also supported by National Institute of Neurological Disorders and Stroke Grant NS48229.

References

Andres FG, Mima T, Schulman AE, Dichgans J, Hallett M, Gerloff C (1999) Functional coupling of human cortical sensorimotor areas during bimanual skill acquisition. Brain 122: 855–870

Astolfi L, Cincotti F, Babiloni C, Carducci F, Basilisco A, Rossini PM, Salinari S, Mattia D, Cerutti S, Ben Dayan D, Ding L, Ni Y, He B, Babiloni F (2005) Estimation of the cortical connectivity by high-resolution EEG and structural equation modeling: Simulations and application to finger tapping data. IEEE Transactions on Biomedical Engineering 52: 757–768

Babiloni F, Cincotti F, Babiloni C, Carducci F, Mattia D, Astolfi L, Basilisco A, Rossini PM, Ding L, Ni Y, Cheng J, Christine K, Sweeney J, He B (2005) Estimation of the cortical functional connectivity with the multimodal integration of high-resolution EEG and fMRI data by directed transfer function. Neuroimage 24: 118–131

Basar, E. (2004) Memory and brain dynamics. Boca Raton, CRC Press.

Beggs, J.M., Klukas, J., Chen, W. Connectivity and dynamics in local cortical networks. This volume.

Binder JR, Rao SM, Hammeke TA, Frost JA, Bandettini PA, Hyde JS (1994) Effects of stimulus rate on signal response during functional magnetic-resonance-imaging of auditory-cortex. Cognitive Brain Research 2: 31–38

Bingham GP, Schmidt RC, Zaal FTJM (1999) Visual perception of the relative phasing of human limb movements. Perception & Psychophysics 61: 246–258

Boecker H, Dagher A, Ceballos-Baumann AO, Passingham RE, Samuel M, Friston KJ, Poline JB, Dettmers C, Conrad B, Brooks DJ (1998) Role of the human rostral supplementary motor area and the basal ganglia in motor sequence control: Investigations with H-2 O-15 PET. Journal of Neurophysiology 79: 1070–1080

Breakspear, M. & Jirsa, V.K. Neural dynamics and brain connectivity. This Volume.

Bressler SL, Kelso JAS (2001) Cortical coordination dynamics and cognition. Trends in Cognitive Neuroscience 5: 26–36

Bressler, S.L. & McIntosh, A.R. The role of neural context in large-scale neurcognitive network applications. This volume.

Brovelli A, Ding MZ, Ledberg A, Chen YH, Nakamura R, Bressler SL (2004) Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality. Proceedings of the National Academy of Sciences of the United States of America 101: 9849–9854

Byblow WD, Chua R, Goodman D (1995) Asymmetries in coupling dynamics of perception and action. Journal of Motor Behavior 27: 123–137

Carson RG, Byblow WD, Abernethy B, Summers JJ (1996) The contribution of inherent and incidental constraints to intentional switching between patterns of bimanual coordination. Human Movement Science 15: 565–589

Carson RG, Riek S (2000) Musculo-skeletal constraints on corticospinal input to upper limb motoneurones during coordinated movements. Human Movement Science 19: 451–474

Carver FW, Fuchs A, Jantzen KJ, Kelso JAS (2002) Spatiotemporal analysis of the neuromagnetic response to rhythmic auditory stimulation: rate dependence and transient to steady-state transition. Clinical Neurophysiology 113: 1921–1931

Case P, Tuller B, Ding MZ, Kelso JaS (1995) Evaluation of a dynamical model of speech-perception. Perception & Psychophysics 57: 977–988

Catalan MJ, Honda M, Weeks RA, Cohen LG, Hallett M (1998) The functional neuroanatomy of simple and complex sequential finger movements: a PET study. Brain 121: 253–264

Chen YQ, Ding MZ, Kelso JAS (2003) Task-related power and coherence changes in neuromagnetic activity during visuomotor coordination. Experimental Brain Research 148: 105–116

Cheyne D, Weinberg H (1989) Neuromagnetic fields accompanying unilateral finger movements–pre-movement and movement–evoked fields. Experimental Brain Research 78: 604–612

Coveney P, Highfield R (1995) Frontiers of complexity: The search for order in a chaotic world. Random House, New York

Crick FH, Koch C (2003) A framework for consciousness. Nature Neuroscience 6: 119–126

Daffertshofer A, Peper CE, Beek PJ (2000) Spectral analysis of event-related encephalographic signals. Physics Letters A 266: 290–302

Dassonville P, Lewis S, Zhu XH, Ugurbil K, Kim SG, Ashe J (1998) Effects of movement predictability on cortical motor activation. Neuroscience Research 32: 65–74

Darvas, F. & Leahy, R.M. Functional imaging of brain activity and connectivity with MEG. This volume.

Debaere F, Wenderoth N, Sunaert S, Van Hecke P, Swinnen SP (2003) Internal vs external generation of movements: differential neural pathways involved in bimanual coordination performed in the presence or absence of augmented visual feedback. Neuroimage 19: 764–776

Debaere F, Wenderoth N, Sunaert S, Van Hecke P, Swinnen SP (2004) Cerebellar and premotor function in bimanual coordination: parametric neural responses to spatiotemporal complexity and cycling frequency. Neuroimage 21: 1416–1427.

DeGuzman, G.C., Tognoli, E., Lagarde, J., Jantzen, K.J., & Kelso, J.A.S. (2005) Effects of the biological relevance of the stimulus in mediating spontaneous visual social coordinaton. *Society for Neuroscience*, Program No. 867.21

Deiber MP, Caldara R, Ibanez V, Hauert CA (2001) Alpha band power changes in unimanual and bimanual sequential movements, and during motor transitions. Clinical Neurophysiology 112: 1419–1435

Deiber MP, Ibanez V, Caldara R, Andrey C, Hauert CA (2005) Programming effectors and coordination in bimanual in-phase mirror finger movements. Cognitive Brain Research 23: 374–386

Dhamala M, Pagnoni G, Wiesenfeld K, Berns GS (2002) Measurements of brain activity complexity for varying mental loads. Physical Review E 65: 1–7

Drake C, Botte MC (1993) Tempo sensitivity in auditory sequences–evidence for a multiple-look model. Perception & Psychophysics 54: 277–286

Edelman GM (2003) Naturalizing consciousness. Proceedings of the National Academy of Sciences of the United States of America 100: 5520–5524

Edelman GM, Tononi G (2000) A Universe of Consciousness. Basic Books, New York

Ehrsson HH, Kuhtz-Buschbeck JP, Forssberg H (2002) Brain regions controlling nonsynergistic versus synergistic movement of the digits: a functional magnetic resonance imaging study. Journal of Neuroscience 22: 5074–5080

Engstrom DA, Kelso JAS, Holroyd T (1996) Reaction-anticipation transitions in human perception-action patterns. Human Movement Science 15: 809–832

Erdler M, Windischberger C, Lanzenberger R, Edward V, Gartus A, Deecke L, Beisteiner R (2001) Dissociation of supplementary motor area and primary motor cortex in human subjects when comparing index and little finger movements with functional magnetic resonance imaging. Neuroscience Letters 313: 5–8

Fox PT, Raichle ME (1984) Stimulus rate dependence of regional cerebral blood-flow in human striate cortex, demonstrated by positron emission tomography. Journal of Neurophysiology 51: 1109–1120

Fraisse P (1982) Rhythm and temp. In: Deutsch D (ed) The psychology of music. Academic Press, New York, pp 149–180

Friston, K.J. (1997). Transients, metastability and neuronal dynamics. Neuroimage, 5, 164–171

Fuchs A. Beamforming and its applications to brain connectivity. This volume.

Fuchs A, Deecke L, Kelso JAS (2000a) Phase transitions in human brain revealed by large SQuID arrays: Response to Daffertshofer, Peper and Beek. Physics Letters A 266: 303–308

Fuchs A, Jirsa VK, Kelso JAS (1999) Traversing scales of brain and behavioral organization II: Analysis and reconstruction. In: Uhl C (ed) Analysis of Neurophysiological Brain Functioning. Springer, Berlin

Fuchs A, Kelso JAS, Haken H (1992) Phase transitions in human brain: Spatial mode dynamics. International Journal of Bifurcation and Chaos 2: 917–939

Fuchs A, Mayville JM, Cheyne D, Weinberg H, Deecke L, Kelso JAS (2000b) Spatiotemporal analysis of neuromagnetic events underlying the emergence of coordinative instabilities. Neuroimage 12: 71–84

Gerloff C, Andres FG (2002) Bimanual coordination and interhemispheric interaction. Acta Psychologica 110: 161–186

Gerloff C, Richard J, Hadley J, Schulman AE, Honda M, Hallett M (1998) Functional coupling and regional activation of human cortical motor areas during simple, internally paced and externally paced finger movements. Brain 121: 1513–1531

Gross J, Kujala J, Hamalainen M, Timmermann L, Schnitzler A, Salmelin R (2001) Dynamic imaging of coherent sources: Studying neural interactions in the human brain. Proceedings of the National Academy of Sciences of the United States of America 98: 694–699

Gross J, Pollok B, Dirks A, Timmermann L, Butz A, Schnitzler A (2005) Task-dependent oscillations during unimanual and bimanual movements in the human primary motor cortex and SMA studied with magnetoencephalography. Neuroimage 26: 91–98

Gross J, Timmermann J, Kujala J, Dirks M, Schmitz F, Salmelin R, Schnitzler A (2002) The neural basis of intermittent motor control in humans. Proceedings of the National Academy of Sciences of the United States of America 99: 2299–2302

Gutschalk A, Mase R, Roth R, Ille N, Rupp A, Hahnel S, Picton TW, Scherg M (1999) Deconvolution of 40 Hz steady-state fields reveals two overlapping source activities of the human auditory cortex. Clinical Neurophysiology 110: 856–868

Haaland KY, Harrington DL, Knight RT (2000) Neural representations of skilled movement. Brain 123: 2306–2313

Haken H (1983) Advances in synergetics. Springer, Berlin

Haken H (1996) Principles of brain function. Springer, Berlin

Haken H, Kelso JAS, Bunz H (1985) A theoretical-model of phase-transitions in human hand movements. Biological Cybernetics 51: 347–356

Hari R, Katila T, Tuomisto T, Varpula T (1982) Interstimulus-interval dependence of the auditory vertex response and its magnetic counterpart. Electroencephalography and Clinical Neurophysiology 53: P71-P72

Harrington DL, Rao SM, Haaland KY, Bobholz JA, Mayer AR, Binder JR, Cox RW (2000) Specialized neural systems underlying representations of sequential movements. Journal of Cognitive Neuroscience 12: 56–77

Haslinger B, Erhard P, Weilke F, Ceballos-Baumann AO, Bartenstein P, von Einsiedel HG, Schwaiger M, Conrad B, Boecker H (2002) The role of lateral premotor-cerebellar-parietal circuits in motor sequence control: a parametric fMRI study. Cognitive Brain Research 13: 159–168

Hock HS, Kelso JAS, Schöner G (1993) Bistability and hysteresis in the organization of apparent motion patterns. Journal of Experimental Psychology-Human Perception and Performance 19: 63–80

Horwitz, B. & Husain, F.T. Simulation frameworks for large-scale brain systems. This volume.

Hummel F, Kirsammer R, Gerloff C (2003) Ipsilateral cortical activation during finger sequences of increasing complexity: representation of movement difficulty or memory load? Clinical Neurophysiology 114: 605–613

Immisch I, Waldvogel D, van Gelderen P, Hallett M (2001) The role of the medial wall and its anatomical variations for bimanual antiphase and in-phase movements. Neuroimage 14: 674–684

Jancke L, Specht K, Mirzazade S, Loose R, Himmelbach M, Lutz K, Shah NJ (1998) A parametric analysis of the 'rate effect' in the sensorimotor cortex: a functional magnetic resonance imaging analysis in human subjects. Neuroscience Letters 252: 37–40

Jantzen KJ, Fuchs A, Mayville JM, Deecke L, Kelso JAS (2001) Neuromagnetic activity in alpha and beta bands reflect learning-induced increases in coordinative stability. Clinical Neurophysiology 112: 1685–1697

Jantzen KJ, Steinberg FL, Kelso JAS (2002) Practice-dependent modulation of neural activity during human sensorimotor coordination: a functional Magnetic Resonance Imaging study. Neuroscience Letters 332: 205–209

Jantzen KJ, Steinberg FL, Kelso JAS (2004) Brain networks underlying human timing behavior are influenced by prior context. Proceedings of the National Academy of Science 101: 6815–6820

Jantzen KJ, Steinberg FL, Kelso JAS (2005) Functional MRI reveals the existence of modality and coordination-dependent timing networks. Neuroimage 25: 1031–1042

Jeka JJ, Kelso JAS (1995) Manipulating symmetry in the coordination dynamics of human movement. Journal of Experimental Psychology-Human Perception and Performance 21: 360–374

Jenkins IH, Passingham RE, Brooks DJ (1997) The effect of movement frequency on cerebral activation: a positron emission tomography study. Journal of Neurological Sciences 151: 195–205

Jirsa VK, Fink P, Foo P, Kelso JAS (2000) Parametric stabilization of biological coordination: a theoretical model. Journal of Biological Physics 26: 85–112

Jirsa VK, Friedrich R, Haken H, Kelso JAS (1994) A theoretical model of phase transitions in the human brain. Biological Cybernetics 71: 27–35

Jirsa VK, Fuchs A, Kelso JAS (1998) Connecting cortical and behavioral dynamics: Bimanual coordination. Neural Computation 10: 2019–2045

Jirsa VK, Haken H (1997) A derivation of a macroscopic field theory of the brain from the quasi-microscopic neural dynamics. Physica D 99: 503–526

Jirsa VK, Jantzen KJ, Fuchs A, Kelso JAS (2002) Spatiotemporal forward solution of the EEG and MEG using network modeling. IEEE Transactions on Medical Imaging 21: 493–504

Jirsa VK, Kelso JAS (2004) (eds.) Coordination dynamics: Issues and trends. Springer

Jirsa, VK and Kelso, JA (2005). The excitator as a minimal model for the coordination dynamics of discrete and rhythmic movement generation. Journal of Motor Behavior 37:35–51

Jirsa VK, Kelso JAS, Fuchs A (1999) Traversing scales of brain and behavioral organization III: Theoretical modeling. In: Uhl C (ed) Analysis of Neurophysiological Brain Functioning. Springer, Berlin

Kawashima R, Inoue K, Sugiura M, Okada K, Ogawa A, Fukuda H (1999) A positron emission tomography study of self-paced finger movements at different frequencies. Neuroscience 92: 107–112

Kelso, J.A.S. (1981). Contrasting perspectives on order and regulation in movement. In A. Baddeley & J. Long (Eds.). *Attention and performance, IX*. Hillsdale, NJ: Erlbaum

Kelso JAS (1984) Phase transitions and critical behavior in human bimanual coordination. American Journal of Physiology 246: 1000–1004

Kelso JAS (1986) Pattern formation in multidegree of freedom speech and limb movement. Experimental Brain Research Supplement 15: 105–128

Kelso JAS (1992) Coordination dynamics of human brain and behavior. Springer Proceedings in Physics 69: 223–234

Kelso JAS (1994) Informational character of self-organized coordination dynamics. Attention and Performance 13: 393–413

Kelso JAS (1995) Dynamic patterns: The self-organization of brain and behavior. MIT Press, Cambridge

Kelso JAS (1997) Relative timing in brain and behavior: Some observations about the generalized motor program and self-organized coordination dynamics. Human Movement Science 16: 453–460

Kelso, J.A.S. & Engstrom, D.A. (2006) The complementary nature. Cambridge, MA: The MIT Press

Kelso, J.A.S. & Fuchs, A. (1995). Self-organizing dynamics of the human brain: Critical instabilities and Sil'nikov chaos. *Chaos, 5,* (1), 64–69

Kelso, J.A.S., Schöner, G., Scholz, J.P. & Haken, H. (1987). Phase-locked modes, phase transitions and component oscillators in coordinated biological motion, *Physica Scripta, 35*, 79–87

Kelso JAS, Bressler SL, Buchanan JJ, de Guzman GG, Ding M, Fuchs A, Holroyd T (1991) Cooperative and critical phenomena in the human brain revealed by multiple SQUIDS. In: Duke D, Pritchard W (eds) Measuring chaos in the human brain. World Scientific, New Jersey, pp 97–112

Kelso JAS, Bressler SL, Buchanan S, Deguzman GC, Ding M, Fuchs a, Holroyd T (1992) A phase-transition in human brain and behavior. Physics Letters A 169: 134–144

Kelso JAS, Delcolle JD, Schöner G (1990) Action-perception as a pattern-formation process. Attention and Performance XIII (M. Jeannerod, ed), Hillsdale, NJ: Erlbaum: 139–169

Kelso JAS, Fink PW, DeLaplain CR, Carson RG (2001) Haptic information stabilizes and destabilizes coordination dynamics. Proceedings of the Royal Society of London Series B-Biological Sciences 268: 1207–1213

Kelso JAS, Fuchs A, Lancaster R, Holroyd T, Cheyne D, Weinberg H (1998) Dynamic cortical activity in the human brain reveals motor equivalence. Nature 392: 814–818

Kelso JAS, Jeka JJ (1992) Symmetry-breaking dynamics of human multilimb coordination. Journal of Experimental Psychology-Human Perception and Performance 18: 645–668

Kelso JAS, Jirsa VK, Fuchs A (1999) Traversing scales of brain and behavioral organization I: Concepts and Experiments. In: Uhl C (ed) Analysis of Neurophysiological Brain Functioning. Springer, Berlin

Kelso JAS, Scholz JP, Schöner G (1988) Dynamics governs switching among patterns of coordination in biological movement. Physics Letters A 134: 8–12

Lagarde J, Kelso JAS (2006) Binding of movement, sound and touch: multimodal coordination dynamics. Experimental Brain Research 173: 673–688

Lewis PA, Wing AM, Pope PA, Praamstra P, Miall RC (2004) Brain activity correlates differentially with increasing temporal complexity of rhythms during initialisation, synchronisation, and continuation phases of paced finger tapping. Neuropsychologia 42: 1301–1312

Lopes da Silva FH, Pfurtscheller G (1999) Basic concepts on EEG synchronization and desynchronization. In: Pfurtscheller G, Lopes da Silva FH (eds) Event-related Desynchronization, vol 6. Elsevier, Amsterdam

Lu ZL, Williamson SJ, Kaufman L (1992) Human auditory primary and association cortex have differing lifetimes for activation traces. Brain Research 572: 236–241

Manganotti P, Gerloff C, Toro C, Katsuta H, Sadato N, Zhuang P, Leocani L, Hallett M (1998) Task-related coherence and task-related spectral power changes during sequential finger movements. Electromyography and Motor Control-Electroencephalography and Clinical Neurophysiology 109: 50–62

Martin, K. A.C. (2006). Where are the switches on this thing? Nature, 440, 1114.

Mayville JM, Bressler SL, Fuchs A, Kelso JAS (1999) Spatiotemporal reorganization of electrical activity in the human brain associated with a timing transition in rhythmic auditory-motor coordination. Experimental Brain Research 127: 371–381

Mayville JM, Fuchs A, Ding MZ, Cheyne D, Deecke L, Kelso JAS (2001) Event-related changes in neuromagnetic activity associated with syncopation and synchronization timing tasks. Human Brain Mapping 14: 65–80

Mayville JM, Fuchs A, Kelso JA (2005) Neuromagnetic motor fields accompanying self-paced rhythmic finger movement at different rates. Experimental Brain Research 166: 190–199

Mayville JM, Jantzen KJ, Fuchs A, Steinberg FL, Kelso JAS (2002) Cortical and subcortical networks underlying syncopated and synchronized coordination revealed using fMRI. Human Brain Mapping 17: 214–229

Meyer-Lindenberg A, Ziemann U, Hajak G, Cohen L, Berman KF (2002) Transitions between dynamical states of differing stability in the human brain. Proceedings of the National Academy of Sciences of the United States of America 99: 10948–10953

Monno A, Temprado JJ, Zanone PG, Laurent M (2002) The interplay of attention and bimanual coordination dynamics. Acta Psychologica 110: 187–211

Nair DG, Purcott KL, Fuchs A, Steinberg F, Kelso JAS (2003) Cortical and cerebellar activity of the human brain during imagined and executed unimanual and bimanual action sequences: a functional MRI study. Cognitive Brain Research 15: 250–260

Neuper C, Pfurtscheller G (2001) Evidence for distinct beta resonance frequencies in human EEG related to specific sensorimotor cortical areas. Clinical Neurophysiology 112: 2084–2097

Nicolis G, Prigogine I (1989) Exploring complexity, an introduction. W.H. Freeman, New York

Nunez PL (1995) Neocortical Dynamics and Human EEG Rhythms. Oxford University Press, New York

Oullier, O., de Guzman, G.C., Jantzen, K.J., & Kelso, J.A.S. (2003). On context dependence of behavioral variability in inter-personal coordination. *International Journal of Computer Science in Sport, 2*, 126–128

Oullier O, Jantzen KJ, Steinberg FL, Kelso JAS (2005) Neural substrates of real and imagined sensorimotor coordination. Cerebral Cortex 15: 975–985

Pantev C, Roberts LE, Elbert T, Ross B, Wienbruch C (1996) Tonotopic organization of the sources of human auditory steady-state responses. Hearing Research 101: 62–74

Pastor MA, Artieda J, Arbizu J, Marti-Climent JM, Penuelas I, Masdeu JC (2002) Activation of human cerebral and cerebellar cortex by auditory stimulation at 40 Hz. Journal of Neuroscience 22: 10501–10506

Pfurtscheller G, Andrew C (1999) Event-related changes of band power and coherence: Methodology and interpretation. Journal of Clinical Neurophysiology 16: 512–519

Pollok B, Gross J, Muller K, Aschersleben G, Schnitzler A (2005) The cerebral oscillatory network associated with auditorily paced finger movements. Neuroimage 24: 646–655

Price C, Wise R, Ramsay S, Friston K, Howard D, Patterson K, Frackowiak R (1992) Regional response differences within the human auditory-cortex when listening to words. Neuroscience Letters 146: 179–182

Rao SM, Bandettini PA, Binder JR, Bobholz JA, Hammeke TA, Stein EA, Hyde JS (1996) Relationship between finger movement rate and functional magnetic resonance signal change in human primary motor cortex. Journal of Cerebral Blood Flow and Metabolism 16: 1250–1254

Rees G, Howseman A, Josephs O, Frith CD, Friston KJ, Frackowiak RSJ, Turner R (1997) Characterizing the relationship between BOLD contrast and regional cerebral blood flow measurements by varying the stimulus presentation rate. Neuroimage 6: 270–278

Ridderikhoff A, Peper CLE, Carson RG, Beek PJ (2004) Effector dynamics of rhythmic wrist activity and its implications for (modeling) bimanual coordination. Human Movement Science 23: 285–313

Riecker A, Wildgruber D, Mathiak K, Grodd W, Ackermann H (2003) Parametric analysis of rate-dependent hemodynamic response functions of cortical and subcortical brain structures during auditorily cued finger tapping: a fMRI study. Neuroimage18: 731–739

Richardson MJ, Marsh KL, Schmidt RC (2005) Effects of visual and verbal interaction on unintentional interpersonal coordination. Journal of Experimental Psychology-Human Perception and Performance 31:62–79.

Rizzolatti G, Luppino G, Matelli M (1998) The organization of the cortical motor system: new concepts. Electroencephalography and Clinical Neurophysiology 106: 283–296

Sabatini U, Chollet F, Rascol O, Celsis P, Rascol a, Lenzi GL, Marcvergnes JP (1993) Effect of side and rate of stimulation on cerebral blood-flow changes in motor areas during finger movements in humans. Journal of Cerebral Blood Flow and Metabolism 13: 639–645

Sadato N, Ibanez V, Campbell G, Deiber MP, LeBihan D, Hallett M (1997) Frequency-dependent changes of regional cerebral blood flow during finger

movements: Functional MRI compared to PET. Journal of Cerebral Blood Flow and Metabolism 17: 670–679
Salmelin R, Hamalainen M, Kajola M, Hari R (1995) Functional segregation of movement-related rhythmic activity in the human brain. Neuroimage 2: 237–243
Sams M, Hari R, Rif J, Knuutila J (1993) The human auditory sensory memory trace persists about 10 sec-neuromagnetic evidence. Journal of Cognitive Neuroscience 5: 363–370
Schmidt RC, Bienvenu M, Fitzpatrick PA, Amazeen PG (1998) A comparison of intra- and interpersonal interlimb coordination: Coordination breakdowns and coupling strength. Journal of Experimental Psychology-Human Perception and Performance 24: 884–900
Schmidt RC, O'Brien B, Sysko R. (1999): Self-organization of between-persons cooperative tasks and possible applications to sport. 30(4):558–579.
Scholz JP, Kelso JAS (1990) Intentional switching between patterns of bimanual coordination is dependent on the intrinsic dynamics of the patterns. Journal of Motor Behavior 22: 98–124
Schöner G, Haken H, Kelso JAS (1986) A stochastic-theory of phase-transitions in human hand movement. Biological Cybernetics 53: 247–257
Schöner G, Kelso JAS (1988) Dynamic pattern generation in behavioral and neural systems. Science 239: 1513–1520
Schöner G, Zanone PG, Kelso JAS (1992) Learning as change of coordination dynamics-theory and experiment. Journal of Motor Behavior 24: 29–48
Serrien DJ, Strens LHA, Oliviero A, Brown P (2002) Repetitive transcranial magnetic stimulation of the supplementary motor area (SMA) degrades bimanual movement control in humans. Neuroscience Letters 328: 89–92
Singer W (1993) Synchronization of cortical activity and its putative role in information-processing and learning. Annual Review of Physiology 55: 349–374
Singer W (1994) The organization of sensory-motor representations in the neocortex-a hypothesis based on temporal coding. Attention and Performance XV 15: 77–107
Singer W (2001) Consciousness and the binding problem. Cajal and Consciousness 929: 123–146
Singh KD, Barnes GR, Hillebrand A, Forde EME, Williams AL (2002) Task-related changes in cortical synchronization are spatially coincident with the hemodynamic response. Neuroimage 16: 103–114
Sporns, O (2004) Complex neural dynamics. In V.K. Jirsa and J.A.S.Kelso (eds) Coordination Dynamics : Issues and Trends, Springer, Heidelberg
Sporns, O & Tononi G. Structural determinants of functional brain dynamics. This volume.
Stephan, KE & Friston, KJ. Models of effective connectivity in neural systems. This volume.
Steyvers M, Etoh S, Sauner D, Levin O, Siebner HR, Swinnen SP, Rothwell JC (2003) High-frequency transcranial magnetic stimulation of the supplementary motor area reduces bimanual coupling during anti-phase but not in-phase movements. Experimental Brain Research 151: 309–317
Swinnen SP (2002) Intermanual coordination: From behavioural principles to neural-network interactions. Nature Reviews Neuroscience 3: 350–361

Swinnen SP, Jardin K, Meulenbroek R, Dounskaia N, HofkensVanDenBrandt M (1997) Egocentric and allocentric constraints in the expression of patterns of interlimb coordination. Journal of Cognitive Neuroscience 9: 348–377

Taniguchi M, Kato A, Fujita N, Hirata M, Tanaka H, Kihara T, Ninomiya H, Hirabuki N, Nakamura H, Robinson SE, Cheyne D, Yoshimine T (2000) Movement-related desynchronization of the cerebral cortex studied with spatially filtered magnetoencephalography. Neuroimage 12: 298–306

Tass PA, Fieseler T, Dammers J, Dolan K, Morosan P, Majtanik M, Boers F, Muren A, Zilles K, Fink GR (2003) Synchronization tomography: A method for three-dimensional localization of phase synchronized neuronal populations in the human brain using magnetoencephalography. Physical Review Letters 90: -

Temprado JJ, Laurent M (2004) Attentional load associated with performing and stabilizing a between-persons coordination of rhythmic limb movements. Acta Psychologica 115: 1–16

Temprado JJ, Monno A, Zanone PG, Kelso JAS (2002) Attentional demands reflect learning-induced alterations of bimanual coordination dynamics. European Journal of Neuroscience 16: 1390–1394

Temprado JJ, Swinnen SP (2005) Dynamics of learning and transfer of muscular and spatial relative phase in bimanual coordination: evidence for abstract directional codes. Experimental Brain Research 160: 180–188

Toma K, Mima T, Matsuoka T, Gerloff C, Ohnishi T, Koshy B, Andres F, Hallett M (2002) Movement rate effect on activation and functional coupling of motor cortical areas. Journal of Neurophysiology 88: 3377–3385

Tononi, G., Sporns, O., Edelman, G.M. (1998) Complexity and coherency: Integrating information in the brain. Trends in Cognitive Science, 2, 474–484.

Tracy JI, Faro SS, Mohammed FB, Pinus AB, Madi SM, Laskas JW (2001) Cerebellar mediation of the complexity of bimanual compared to unimanual movements. Neurology 57: 1862–1869

Tschacher W, Dauwalder JP (eds) (2003) Embodiment and coordination dynamics, studies of nonlinear phenomenon in life science. World Scientific, New Jersey

Tuller B, Case P, Ding MZ, Kelso JaS (1994) The nonlinear dynamics of speech categorization. Journal of Experimental Psychology-Human Perception and Performance 20: 3–16

Turvey MT (2004) Impredicativity, dynamics and the perception-action divide. In: Jirsa VK, Kelso JAS (eds) Coordination Dynamics: Issues and Trends. Springer, Berlin

Ullen F, Bengtsson SL (2003) Independent processing of the temporal and ordinal structure of movement sequences. Journal of Neurophysiology 90: 3725–3735

VanOostende S, VanHecke P, Sunaert S, Nuttin B, Marchal G (1997) FMRI studies of the supplementary motor area and the premotor cortex. Neuroimage 6: 181–190

Varela F, Lachaux JP, Rodriguez E, Martinerie J (2001) The brainweb: Phase synchronization and large-scale integration. Nature Reviews Neuroscience 2: 229–239

Wallenstein GV, Kelso JAS, Bressler SL (1995) Phase-transitions in spatiotemporal patterns of brain activity and behavior. Physica D, 84: 626–634

Wexler BE, Fulbright RK, Lacadie CM, Skudlarski P, Kelz MB, Constable RT, Gore JC (1997) An fMRI study of the human cortical motor system response to increasing functional demands. Magnetic Resonance Imaging 15: 385–396

Zaal FTJM, Bingham GP, Schmidt RC (2000). Visual perception of mean relative phase and phase variability. Journal of Experimental Psychology-Human Perception and Performance 26: 1209–1220.

Zanone PG, Kelso JAS (1991) Experimental studies of behavioral attractors and their evolution with learning. In: Requin J, Stelmach GE (eds) Tutorials in motor neurosciences. Kluwer, Dordrecht, pp 121–133

Zanone PG, Kelso JAS (1992) Evolution of behavioral attractors with learning-nonequilibrium phase-transitions. Journal of Experimental Psychology-Human Perception and Performance 18: 403–421

Zanone PG, Kelso JAS (1997) Coordination dynamics of learning and transfer: Collective and component levels. Journal of Experimental Psychology-Human Perception and Performance 23: 1454–1480

Zanone PG, Monno A, Temprado JJ, Laurent M (2001) Shared dynamics of attentional cost and pattern stability. Human Movement Science, 20: 765–789

Maturation of Structural and Functional Connectivity in the Human Brain

Tomáš Paus

Brain & Body Centre, University of Nottingham, United Kingdom and Montreal Neurological Institute, McGill University, Canada

Smooth flow of neural impulses throughout the brain allows for information to be integrated across the many spatially segregated and functionally specialized modules. Structural and functional maturation of neural pathways connecting individual brain regions is therefore a condition *sine qua non* for the successful development of cognitive, motor and sensory functions from infancy, through childhood and adolescence, and into adulthood.

Typically, region A is said to be connected with region B only if neurons A possess synaptic connections with neurons B. Techniques available today for the study of neural connectivity in the human brain do not provide such a level of spatial neuron-to-neuron specificity, however. With the exception of *post-mortem* studies of short-range cortical connectivity with the carbocyanin tracer DiI (e.g. Tardif and Clarke 2001), most of the current research focuses on *in vivo* studies of structural and functional connectivity at the macroscopic level. This work is carried out with a variety of brain mapping techniques, including structural and functional magnetic resonance imaging (MRI), positron emission tomography (PET), transcranial magnetic stimulation (TMS), electroencephalography (EEG) and magneto-encephalography (MEG). In this context, the term "connectivity" refers either to the structural properties of white matter and major fibre tracts (structural connectivity) or to the statistical relationship in neural activity recorded simultaneously in a number of spatially distinct regions (functional connectivity). Under certain circumstances, we can also evaluate how one region influences another (effective connectivity). In this chapter, I will review the current knowledge regarding maturation of structural, functional and effective connectivity as assessed *in vivo* with some of the above brain mapping tools.

1 Structural Connectivity

The initial developmental studies of structural connectivity in newborns and infants assessed changes in magnetic resonance (MR) relaxation times and in the qualitative appearance of T1- and T2-weighted images, focusing on the

timing of the initial reversal of grey-white differentiation and on the temporal order in which different brain structure first acquire a "myelinated" appearance (reviewed in Paus et al. 2001). Studies carried out in children and adolescents allowed investigators to evaluate age-related changes in the overall and regional volumes in white matter. To date, one of the largest datasets of MR images obtained in typically developing children and adolescents has been acquired at the Child Psychiatry Branch of the National Institute of Mental Health (NIMH) in the U.S.A. A combination of cross-sectional (161 subjects) and longitudinal (329 scans) data is available (Giedd 2004). In this cohort, Giedd and colleagues observed a steady increase in the overall volume of white matter throughout the studied age-range of 4 to 21 years (Giedd et al. 1999a). This finding is consistent with previous observations (Pfefferbaum et al. 1994 [n = 88, 1 to 30 years]; Reiss et al. 1996 [n = 85, 5 to 17 years]). In a subsample of the NIMH cohort (n = 111), Paus et al. (1999) have documented an increase in white-matter density in the internal capsule and the left arcuate fasciculus, the latter containing fibres presumably connecting the anterior (Broca) and posterior (Wernicke) speech regions. Another example of regional variation in white-mater maturation is a region-specific growth of the corpus callosum; both cross-sectional and longitudinal analyses revealed continuous age-related changes in its posterior but not the anterior section (Giedd et al. 1996, 1999b, Pujol et al. 1993). In a recent study, Blanton et al. (2004) documented significant gender differences in white matter of the left inferior frontal gyrus, a region consisting of the pars opercularis, triangularis and orbitalis and, presumably, containing speech-related areas: boys (n=25, 6 to 17 years) but not girls (n=21, 6 to 15 years) showed a linear age-related increase in the WM volume in this region. This regional age-gender interaction is consistent with the steeper age-related increases in the overall white-matter volume in boys, compared with girls (Giedd et al. 1999a).

With a few exceptions (corpus callosum), the above studies suffer from the absence of natural anatomic boundaries in white matter. Note, however, that voxel-wise analyses are capable of detecting regional effects in poorly defined white-matter structures, such as the internal capsule and the putative arcuate fasciculus (Paus et al. 1999). Nonetheless, the ability to define precisely major fibre tracts with diffusion tensor imaging (DTI) facilitates further the regional evaluation of structural properties of distinct white-matter systems. This imaging technique has been employed to assess age-related changes in magnitude (apparent diffusion coefficient, ADC) and directionality (fractional anisotropy, FA) of the diffusion of water in the human brain during childhood and adolescence. Overall, DTI-based studies reveal age-related decreases in ADC and increases in FA in a number of WM regions, many of which are identical to those revealed by the above MR studies (Klingberg et al. 1999, Schmithorts et al. 2002, Snook et al. 2005); changes in FA are interpreted as reflecting an increase in density of fibres and/or increase in myelination. It is of note that FA in the white matter of frontal lobes appears to

correlate with IQ (Schmithorst et al. 2005) while FA in the left temporo-parietal white matter correlates with reading skills (Beaulieu et al. 2005, Deutsch et al. 2005). Interpretations of DTI-based observations should be made with caution, however. As much as this technique allows investigators to visualize distinct fibre tracts, it does not speak to the point-to-point connectivity between spatially distinct brain regions; individual axons may join and leave fibre tracts undetected. To some extent, this limitation may be overcome in the future by further refinements of fibre tractography (e.g. Lee, Mori et al. 2005, Johansen-Berg et al. 2005). It is also unclear which of the cellular properties affect water diffusion the most. Movement of the water molecules in the intracellular and extracellular space can be affected by structural properties of the axon (e.g. myelin sheath, microtubules, neurofilaments) as well as by the indirect effects of neighbouring cells on the size and shape of the extracellular space.

Overall, it is clear that, at the macroscopic level, structural connectivity of the human brain continues to mature during childhood and adolescence. Future studies will clarify which aspects of structural connectivity are most dynamic, and whether or not experience plays a role in shaping structural connectivity in a similar manner as it does in the case of grey matter.

2 Functional Connectivity

Functional connectivity can be defined operationally as the extent of correlation in brain activity measured across a number of spatially distinct brain regions (e.g. Friston 1994, Horwitz 2003, Sporns et al. 2004). When discussing various approaches to the study of functional connectivity in the *adult* human brain, Horwitz (2003) pointed out that conclusions reached by different investigators regarding the presence or absence of functional connectivity between a set of brain regions depend on the type of measurement (e.g. functional MRI, EEG, MEG), type of analysis (e.g. correlation, structural equation modelling) and, most importantly, the state of the subject during the recording of brain activity (rest, type of stimulation/task). This is, or will be, very much the case in studies of the developing brain. At the moment, however, the majority of investigators have employed only two approaches: (1) coherence of EEG signals; and (2) inter-regional correlations in fMRI time series.

Coherence is analogous to the squared correlation in the frequency domain between two EEG signals (time series) measured simultaneously at different scalp locations (Nunez 1981). In most previous EEG studies of brain development, EEG was recorded *at rest* with a relatively small number of electrodes (< 20) placed over the left and right hemispheres according to the international 10–20 system, and analyzed in four frequency bands (delta, theta, alpha and beta). When interpreting differences in EEG coherence between electrodes with different physical distance within (e.g. F3-P3 vs. F3-O1) or between (O1-O2 vs. T5-T6) hemispheres, two important factors come to play.

Volume conduction inflates coherence at short (< 10 cm) distances, while an increasing phase difference reduces coherence at large (> 15 cm) distances (see e.g. Nunez et al. 1997, Srinivasan et al. 1998, Barry et al. 2005). Thatcher et al. (1986) mapped the physical distinction between "short" and "large" distances between scalp electrodes on to an anatomical one, namely the Braitenberg's (1978) two-compartment model of axonal systems in the cerebral cortex. According to Braitenberg, compartment "A" is composed of the basal dendrites that receive input from the axon collaterals given off by the neighbouring pyramidal cells, while compartment "B" consists of the apical dendrites of pyramidal cells that receive input from "large-distance" cortico-cortical projections. The subsequent empirical work on the development of EEG coherence has provided the following observations.

Using EEG recorded with 19 electrodes in 577 children and adolescents (age from 2 months to 26 years), Thatcher et al. (1987) described increases in EEG coherence that took place over different brain regions at different periods of development. For example, fronto-occipital coherence increased in the left hemisphere between 4 to 6 years, while fronto-temporal coherence increased in the right hemisphere from 8 to 10 years (Thatcher et al. 1987). In another sample (n = 253, 6 months to 7 years), Thatcher et al. (1992) observed rapid changes ("growth spurts") in EEG coherence that suggested opposite developmental changes taking place in the two hemispheres: increases in large-distance and short-distance coherence, respectively, in the left and right hemispheres. More recently, Barry et al. (2004), measured EEG coherence (18 electrodes) in 80 children between 8 and 12 years of age (8 boys and 8 girls in each 1-year age-band) and reported interesting age- and gender effects in the maturation of large-distance coherence. For example, intra-hemispheric coherence increased with age in all but the beta band, boys showed an overall higher coherence in the higher EEG bands and, unlike girls, a continued increase in alpha coherence at older ages. Note that these gender differences are consistent with the more pronounced volumetric increase of white matter in boys, as compared with girls, as described in the structural MR studies (see above). They also observed age-related changes in inter-hemispheric coherence; robust age-related changes in EEG coherence took place between the (left and right) temporal but not the (left and right) frontal electrodes; this finding is consistent with the continuing growth of the splenium of the corpus callosum (e.g. Giedd et al. 1996). Using a 128 channel Geodesic Sensor Net (mean nearest-neighbour electrode separation of 2.7 cm), Srinivasan (1999) computed EEG coherence in alpha band for every possible pair of electrodes ($\sim 6,000$ pairs) and compared the values obtained in children (n = 20, 6 to 11 years) and young adults (n = 23, 18–23 years). For electrode pairs separated by more than 15 cm (i.e. large-distance coherence), he found higher intra-hemispheric coherence in adults, as compared with children, in both hemispheres; these "large-distance" pairs are between electrodes placed over the frontal and occipital cortex. Srinivasan speculated that such an increase in the EEG coherence between distant anterior and posterior sites in the same

hemisphere might be related to the structural maturation of long-range fibre tracts such as the arcuate fasciculus, which we did describe in our structural MR studies (Paus et al. 1999, Watkins et al. 2002). In addition, Srinivasan observed more subtle differences between children and adults in large-distance inter-hemispheric coherence, but only for the anterior electrodes. The latter finding is at odds, however, with those of Barry et al. (2004) and with the clear age-related differences in the size of the splenium of the corpus callosum (see above). Finally, van Baal and colleagues (2001) examined heritability of EEG coherence in theta band and its possible (longitudinal) changes over an 18-month period during childhood (from 5.3 years to 6.8 years). In a total of 152 monozygotic and dizogotic twin pairs, they recorded resting EEG at 10 scalp locations and computed genetic and environmental contributions to short- and large-distance intra-hemispheric coherence. The overall heritabilities were moderate to high for all coherences (average 58%). Heritability of short-distance coherence recorded over the frontal and occipital lobes appeared to decrease and increase with age, respectively. At the second time point (age 7), new genetic factors were found for some of the large-distance coherences, such as the one between prefrontal and parietal electrodes.

Overall, the existing developmental studies of EEG coherence show clearly the potential of this approach for the assessment of functional connectivity in typically developing children and adolescents. One of the main advantages of EEG is its high temporal resolution; high coherence indicates similar values of phase angle between two channels from one EEG epoch to the next, with the epoch duration being typically about two seconds and the recording lasting for a few minutes. But given the absence of experimentally controlled stimulation in most developmental studies of coherence, it is likely that functional connectivity assessed by coherence may be limited only to the neural circuits that display spontaneous oscillations at rest and that are, at the same time, inter-connected structurally, whether directly (e.g. cortico-cortical) or indirectly (e.g. cortico-subcortical-cortical). Nonetheless, assessment of EEG coherence, in particular in combination with structural MRI and/or DTI obtained in the same individual, will likely continue to further our understanding of physiological processes underlying the maturation of functional connectivity in humans.

Today, functional fMRI represents the most common approach for mapping age-related changes in brain activity in childhood and adolescence; its high spatial resolution makes this technique ideal for determining *where* in the brain changes in neural activity occur. It is important to note, however, that the relationship between fMRI signal and brain activity is not straightforward. Changes in local hemodynamics most likely reflect the sum of excitatory postsynaptic inputs in the sample of scanned tissue (Logothetis et al. 2001). Inhibitory neurotransmission may lead to decreases in CBF indirectly, through its presynaptic effects on postsynaptic excitation (Mathiesen et al. 1998). Interpretation of the BOLD signal aside, the use of fMRI in studies of children and adolescents presents many specific challenges of both technical (e.g. head

motion and its correction) and conceptual (e.g. age vs. performance) nature (reviewed in Davidson et al., 2003).

Over the past 10 years, a number of fMRI studies examined neural correlates of developmental changes in various cognitive functions, such as working memory, behavioural inhibition, language or face processing. This work has been done not only in children and adolescents (reviewed in Paus 2005a, Durston and Casey 2005) but also in infants (e.g. Dehaene-Lambertz et al. 2002). But only a handful of investigators included assessment of functional connectivity in their analyses. In a study of memory encoding, Menon et al. (2005) observed age-related (n = 15, 11 to 19 years) *decrease* in the fMRI signal in the left medial temporal lobe while subjects viewed a series of novel photographs of natural outdoor scenes, as compared with viewing the same scene (control condition). They then used voxel-wise regression analysis to identify brain regions that showed correlation in the fMRI signal with that measured in two subregions of the left medial temporal lobe, namely the hippocampus and the entorhinal cortex. This analysis revealed an age-related *increase* in the correlation between the left entorhinal cortex and the left dorsolateral prefrontal cortex. This work nicely illustrates the importance of including the analysis of functional connectivity in developmental studies: although the fMRI signal decreased in one of the memory-relevant structures (entorhinal cortex), the hypothesized interaction between this structure and other brain regions (prefrontal cortex) actually increased. In a different developmental study, Schmithorst and Holland (2006) investigated the relationship between intelligence and functional connectivity in a large sample of typically developing children and adolescents (n=323, 5 to 18 years). They measured fMRI signal during a task requiring the child to generate silently appropriate verbs in response to hearing nouns (every 5 sec). After identifying brain regions engaged during this task, as compared with simply tapping fingers in response to a warble tone, they correlated fMRI signal in all such voxels with the subjects' full-scale intelligence quotient (FSIQ). This analysis revealed positive fMRI-FSIQ correlation in five regions of the left hemisphere: Broca's area, middle temporal gyrus, the anterior cingulate, the precuneus and the medial frontal gyrus (putative supplementary motor area). Next, the authors computed the connectivity coefficient, defined as a weighted sum of the pairwise covariances between these regions. Using the connectivity coefficient as a measure of functional connectivity, Schmithorst and Holland found the following age and gender effects vis-à-vis the relationship between intelligence and connectivity. In boys, functional connectivity appeared to increase as a function of intelligence between age 5 and 9 years; no such relationship was present in older boys (10 to 12 years), and a negative correlation was observed in the oldest boys (13 and 18 years). In girls, on the other hand, no relationship was found in younger girls (age 5 to 13 years) but a strong positive correlation between functional connectivity and intelligence was clearly present in older girls (13 to 18 years). It is of note that the above effects were found in the time series measured during both the verb-generation and the control

task, while some other effects were only found when the control time series was subtracted from the task one. As the authors point out in their discussion, the above gender differences in the relationship between intelligence and functional connectivity are consistent with some structural findings in adults, such as significant correlations between intelligence and regional white-matter volumes in women but grey-matter volumes in men (Haier et al., 2005).

Overall, studies of functional connectivity with fMRI are in their infancy. The two examples given above illustrate the power of this approach but also indicate that large numbers of subjects of both genders in different age groups may be necessary to reach valid conclusions. This is clearly a challenge; acquisition of high-quality fMRI datasets in children is not easy. As illustrated elsewhere (e.g. Schlaggar et al. 2002), possible differences in performance between subjects of different ages (and gender) must also be taken into account when interpreting fMRI data. Finally, acquiring high-resolution structural MRI and/or DTI in the same individuals would greatly facilitate the identification of possible similarities, and differences, between functional and structural connectivity in the same sample.

3 Effective Connectivity

Effective connectivity attempts to describe causal effects exerted by one brain region onto another (Friston 1994). As pointed out by Sporns et al. (2004), effective connectivity can be inferred through a perturbation or through the observation of temporal order of neural events (Granger causality; Granger 1969). At least theoretically, the latter approach is possible in the case of electrophysiological signals. It is still unclear, however, whether relatively short (a few milliseconds) delays in monosynaptic pathways can be discerned using EEG or MEG measures, which are based on a spatially integrated response of a large population of neurons. The perturbation approach seems to be more widely used at the moment; it typically involves stimulation of one cortical region with TMS while brain activity is recorded simultaneously with EEG, PET, or fMRI (reviewed in Komssi and Kahkonen 2006, Paus 2005b, 2002). But the combination of brain stimulation with brain imaging is a technically challenging approach that places high demands on both the experimenter and the subject. It is therefore not surprising that only one developmental study has so far combined TMS with brain mapping, namely with EEG (Bender et al. 2005). Rather than studying connectivity, however, these authors focused on local EEG effects of single-pulse TMS applied over the primary motor cortex in healthy children (n = 17, 6 to 10 years).

Nonetheless, there is a relative wealth of data about the development of effective connectivity of the motor cortex, both with the spinal cord and the contralateral motor cortex (see Garvey and Gilbert 2004 for a review of the use of TMS in children). Applying a single pulse of TMS over the motor cortex elicits a muscle evoked-potential (MEP) in the contralateral (hand)

muscles. One can calculate the time it takes for such a TMS-induced activation to reach the (cervical) spinal cord; the so called central motor conduction time (CMCT) is computed by subtracting the latency of the MEP elicited by stimulating cervical motor roots from that elicited by stimulating the motor cortex. Age-related shortening of the CMCT has been reported in several studies of healthy children and adolescents (Nezu et al. 1997, n = 46, 1 to 14 years; Muller et al. 1994, n = 91, 2 months to 13 years; Eyre et al. 1991, n = 308, newborn to 55 years). These studies clearly show that motor cortex possesses effective connectivity with the spinal cord from an early age but, most importantly, that the speed with which signals are transmitted along this pathway increases as a function of age. This change is fast during the first two years of life and slow, but present, until late adolescence. The two phases likely correspond, respectively, to the initial robust myelination of the corticospinal tract during infancy (Yakovlev and Lecours 1967) and its continuing structural maturation during childhood and adolescence (Paus et al. 1999). A handful of developmental studies also address the question of effective connectivity between the left and right primary motor cortex. The experimental approach used in these studies takes advantage of the phenomenon called transcallosal inhibition: in adults, a single pulse of TMS applied over, say, left motor cortex not only elicits an MEP in the muscles of the right hand but it also suppresses muscle activity, which is maintained by a voluntary contraction of the muscles, in the left hand. It is likely that this suppression is mediated by inter-hemispheric cortico-cortical connections, from the left to the right motor cortex, passing through the posterior half of the trunk of the corpus callosum (Meyer et al. 1995, 1998). It appears that transcallosal inhibition cannot be elicited in young children (Heinen et al. 1998, n = 7, 4.2 to 5.7 years), emerging only between the age of 6 and 10 years (Garvey et al. 2003, n = 34, 6.4 to 13.9 years). After the age of 10 years, it is present in most healthy children (Garvey et al. 2003; Heinen et al. 1999, n = 4, 10 to 15 years). Interesting hemispheric asymmetries exist in effective connectivity between the two motor cortices in that transcallosal inhibition of the dominant motor cortex develops earlier than that of the non-dominant motor cortex (Garvey et al. 2003). Even when present, latency of the transcallosal inhibition is longer in younger (6 to 9 years) compared with older (10 to 14 years) children (Garvey et al. 2003). Thus, despite the presence of a well-developed corpus callosum, it seems that effective connectivity between the primary motor cortices of the two hemispheres does not develop until late childhood. It is possible that subtle age-related changes in structural connectivity (measured, for example, with DTI) may explain the emergence of inter-hemispheric effective connectivity and, in particular, its shorter latency in older children. But it is equally, if not more, likely that suppression of ongoing activity in the motor cortex by the transcallosal input also requires a mature intracortical (inhibitory) circuitry.

Overall, the handful of TMS-based studies of the developing motor system suggests that the primary motor cortex possesses effective connectivity with the spinal cord from an early age (< 1 year) whereas its effective connectivity with the contralateral motor cortex develops only later (between 6 and 10 years). In both cases, the speed of signal transmission along a given pathway continues to increase with the child's age.

4 Conclusions

Over the last 20 years, methodological advancements in *in vivo* measurement of neural connectivity have opened up new avenues for studies of brain maturation in infancy, childhood and adolescence. In the mid 1980's, initial studies of EEG coherence focused on developmental trajectories in local (short-range) and distal (long-range) coherence, the latter most likely reflecting functional connectivity in anatomically connected networks oscillating spontaneously at a certain frequency. The mid and late 1990's brought a wealth of data documenting age-related changes in the global and regional volumes of white matter; we have learned that at least some of the volumes continue to increase into the third decade of life and that significant gender differences exist in the size and location of these maturational effects. The new century saw further developments in the application of DTI to the refinement of developmental studies of structural connectivity and the first attempts to investigate functional connectivity by inter-regional correlation of fMRI timeseries. Finally, maturational studies of effective connectivity to date remain limited to a TMS-based assessment of the motor cortex and its connectivity with the spinal cord and the contralateral motor cortex.

Where do we go from here? It is likely that the current acquisition and analysis methods will continue to improve and increase, in turn, the spatial and temporal resolution of our datasets and the statistical power necessary for revealing subtle age-related variations in structural and functional connectivity. It is also clear that an increasing number of studies will attempt to collect multimodal datasets (e.g. EEG and structural MR) in order to investigate similarities and differences in structural and functional connectivity, thus explaining age-related variance at different levels. The main challenge here will be that of collecting large enough samples representing different age groups and both genders. But, as always, the new breakthroughs will come from discoveries of new measurement techniques allowing us, for example, to study intra-cortical (structural and functional) connectivity within a cortical area or to perturb a given cortical region in a highly selective (and non-invasive) manner. It is without a doubt that future studies of structural and functional connectivity will provide important insights regarding the environmental and genetic mechanisms shaping the human brain.

Acknowledgements

The author's work in supported by the Canadian Institutes of Health Research, the Royal Society (U.K.), and the National Institutes of Health (U.S.A.).

References

Barry RJ, Clarke AR, McCarthy R, Selikowitz M, Johnstone SJ, Rushby JA. Age and gender effects in EEG coherence: I. Developmental trends in normal children. Clin Neurophysiol. 2004 Oct;115(10):2252–8.

Barry RJ, Clarke AR, McCarthy R, Selikowitz M. Adjusting EEG coherence for inter-electrode distance effects: an exploration in normal children. Int J Psychophysiol. 2005 Mar;55(3):313–21.

Bender S, Basseler K, Sebastian I, Resch F, Kammer T, Oelkers-Ax R, Weisbrod M. Electroencephalographic response to transcranial magnetic stimulation in children: Evidence for giant inhibitory potentials. Ann Neurol. 2005 Jul;58(1):58–67.

Blanton RE, Levitt JG, Peterson JR, Fadale D, Sporty ML, Lee M, To D, Mormino EC, Thompson PM, McCracken JT, Toga AW. Gender differences in the left inferior frontal gyrus in normal children. Neuroimage. 22:626–36, 2004.

Braitenberg V. Brain size and number of neurons: an exercise in synthetic neuroanatomy. J Comput Neurosci. 10:71–7, 2001.

Brown TT, Lugar HM, Coalson RS, Miezin FM, Petersen SE, Schlaggar BL. Developmental Changes in Human Cerebral Functional Organization for Word Generation. Cereb Cortex. Aug 5, 2004.

Beaulieu C, Plewes C, Paulson LA, Roy D, Snook L, Concha L, Phillips L. Imaging brain connectivity in children with diverse reading ability. Neuroimage. 2005 May 1;25(4):1266–71.

Deutsch GK, Dougherty RF, Bammer R, Siok WT, Gabrieli JD, Wandell B. Children's reading performance is correlated with white matter structure measured by diffusion tensor imaging. Cortex. 2005 Jun;41(3):354–63.

Davidson MC, Thomas KM, Casey BJ. Imaging the developing brain with fMRI. Ment Retard Dev Disabil Res Rev. 9:161–7, 2003.

Dehaene-Lambertz G, Dehaene S, Hertz-Pannier L. Functional neuroimaging of speech perception in infants. Science. 2002 Dec 6;298(5600):2013–5.

Durston S, Casey BJ. What have we learned about cognitive development from neuroimaging? Neuropsychologia. 2005 Nov 19

Eyre JA, Miller S, Ramesh V. Constancy of central conduction delays during development in man: investigation of motor and somatosensory pathways. J Physiol. 1991 Mar;434:441–52.

Friston KJ. Functional and effective connectivity in neuroimaging: A synthesis. (1994) Human Brain Mapping 2, 56–78.

Garvey MA, Gilbert DL. Transcranial magnetic stimulation in children. Eur J Paediatr Neurol. 2004;8(1):7–19

Garvey MA, Ziemann U, Bartko JJ, Denckla MB, Barker CA, Wassermann EM. Cortical correlates of neuromotor development in healthy children. Clin Neurophysiol. 2003 Sep;114(9):1662–70.

Giedd JN. Structural magnetic resonance imaging of the adolescent brain. Ann N Y Acad Sci. 1021:77–85, 2004.

Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, Paus T, Evans AC and Rapoport JL. Brain development during childhood and adolescence: a longitudinal MRI study. Nature Neuroscience. 2: 861–863, 1999a.

Giedd JN, Blumenthal J, Jeffries NO, Rajapakse JC, Vaituzis AC, Liu H, Berry YC, Tobin M, Nelson J and Castellanos FX. Development of the human corpus callosum during childhood and adolescence: a longitudinal MRI study. Progress in Neuro-Psychopharmacology & Biological Psychiatry. 23: 571–588, 1999b.

Giedd JN, Rumsey JM, Castellanos FX, Rajapakse JC, Kaysen D, Vaituzis AC, Vauss YC, Hamburger SD and Rapoport JL. A quantitative MRI study of the corpus callosum in children and adolescents. Developmental Brain Research. 91: 274–280, 1996.

Granger CWJ. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, Econometrica, 37:424–438, 1969.

Haier RJ, Jung RE, Yeo RA, Head K, Alkire MT. The neuroanatomy of general intelligence: sex matters. Neuroimage. 2005 Mar;25(1):320–7.

Horwitz B. The elusive concept of brain connectivity. Neuroimage. 2003 Jun;19 (2 Pt 1):466–70.

Johansen-Berg H, Behrens TE, Sillery E, Ciccarelli O, Thompson AJ, Smith SM, Matthews PM. Functional-anatomical validation and individual variation of diffusion tractography-based segmentation of the human thalamus. Cereb Cortex. 2005 Jan;15(1):31–9.

Klingberg T, Vaidya CJ, Gabrieli JDE, Moseley ME and Hedehus M. Myelination and organization of the frontal white matter in children: a diffusion tensor MRI study. NeuroReport. 10: 2817–2821, 1999.

Komssi S, Kahkonen S. The novelty value of the combined use of electroencephalography and transcranial magnetic stimulation for neuroscience research. Brain Res Brain Res Rev. 2006 Mar 15;

Lee SK, Kim DI, Kim J, Kim DJ, Kim HD, Kim DS, Mori S. Diffusion-tensor MR imaging and fiber tractography: a new method of describing aberrant fiber connections in developmental CNS anomalies. Radiographics. 2005 Jan-Feb;25(1):53–65.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. Neurophysiological investigation of the basis of the fMRI signal. Nature 412:150–157, 2001.

Mathiesen C, Caesar K, Akgoren N, Lauritzen M. Modification of activity-dependent increases of cerebral blood flow by excitatory synaptic activity and spikes in rat cerebellar cortex. J Physiol 512:555–566, 1998.

Menon V, Boyett-Anderson JM, Reiss AL. Maturation of medial temporal lobe response and connectivity during memory encoding. Brain Res Cogn Brain Res. 2005 Sep;25(1):379–85.

Meyer BU, Roricht S, Grafin von Einsiedel H, Kruggel F, Weindl A. Inhibitory and excitatory interhemispheric transfers between motor cortical areas in normal humans and patients with abnormalities of the corpus callosum. Brain. 1995 Apr;118 (Pt 2):429–40.

Meyer BU, Roricht S, Woiciechowsky C. Topography of fibers in the human corpus callosum mediating interhemispheric inhibition between the motor cortices. Ann Neurol. 1998 Mar;43(3):360–9.

Muller K, Ebner B, Homberg V. Maturation of fastest afferent and efferent central and peripheral pathways: no evidence for a constancy of central conduction delays. Neurosci Lett. 1994 Jan 17;166(1):9–12.

Nezu A, Kimura S, Uehara S, Kobayashi T, Tanaka M, Saito K. Magnetic stimulation of motor cortex in children: maturity of corticospinal pathway and problem of clinical application. Brain Dev. 1997 Apr;19(3):176–80.

Nunez PL Electric Fields of the Brain, The Neurophysics of EEG. New York: Oxford University, 1981.

Nunez PL, Srinivasan R, Westdorp AF, Wijesinghe RS, Tucker DM, Silberstein RB, Cadusch PJ. EEG coherency. I: Statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. Electroencephalogr Clin Neurophysiol. 1997 Nov;103(5):499–515.

Olesen PJ, Nagy Z, Westerberg H, Klingberg T. Combined analysis of DTI and fMRI data reveals a joint maturation of white and grey matter in a fronto-parietal network. Brain Res Cogn Brain Res. 18:48–57, 2003.

Paus, T. Combination of Transcranial Magnetic Stimulation with Brain Imaging. In: J. Mazziotta, A. Toga (Eds). Brain Mapping: The Methods. Second Edition Academic Press, pp. 691–705, 2002.

Paus T. Mapping brain maturation and cognitive development during adolescence. Trends in Cognitive Sciences 9:60–68, 2005a.

Paus T. Inferring causality in brain images: a perturbation approach. Philosophical Transactions of the Royal Society, Series B 360:1109–1114, 2005b

Paus T, Zijdenbos A, Worsley K, Collins DL, Blumenthal J, Giedd JN, Rapoport JL and Evans AC. Structural maturation of neural pathways in children and adolescents: in vivo study. Science. 283: 1908–1911, 1999.

Paus T, Collins DL, Evans AC, Leonard G, Pike B, Zijdenbos A. Maturation of white matter in the human brain: a review of magnetic-resonance studies. Brain Research Bulletin 54:255–266, 2001.

Pfefferbaum A, Mathalon DH, Sullivan EV, Rawles JM, Zipursky RB, Lim KO. A quantitative magnetic resonance imaging study of changes in brain morphology from infancy to late adulthood. Archives of Neurology. 51: 874–887, 1994.

Pujol J, Vendrell P, Junque C, Marti-Vilalta JL and Capdevila A. When does human brain development end? Evidence of corpus callosum growth up to adulthood. Annals of Neurology. 34: 71–75, 1993.

Reiss AL, Abrams MT, Singer HS, Ross JL and Denckla MB. Brain development, gender and IQ in children. A volumetric imaging study. Brain. 119: 1763–1774, 1996.

Schlaggar BL, Brown TT, Lugar HM, Visscher KM, Miezin FM, Petersen SE. Functional neuroanatomical differences between adults and school-age children in the processing of single words. Science. 296:1476–9, 2002.

Schmithorst VJ, Holland SK. Functional MRI evidence for disparate developmental processes underlying intelligence in boys and girls. Neuroimage. 2006 Mar 13;

Schmithorst VJ, Wilke M, Dardzinski BJ, Holland SK. Cognitive functions correlate with white matter architecture in a normal pediatric population: A diffusion tensor MRI study. Hum Brain Mapp. 2005 Apr 27;

Schmithorst VJ, Wilke M, Dardzinski BJ, Holland SK. Correlation of white matter diffusivity and anisotropy with age during childhood and adolescence: a cross-sectional diffusion-tensor MR imaging study. Radiology. 2002 Jan;222(1):212–8.

Snook L, Paulson LA, Roy D, Phillips L, Beaulieu C. Diffusion tensor imaging of neurodevelopment in children and young adults. Neuroimage. 2005 Jul 15;26(4):1164–73.

Sporns O, Chialvo DR, Kaiser M, Hilgetag CC.Organization, development and function of complex brain networks. Trends Cogn Sci. 2004 Sep;8(9):418–25.

Srinivasan R. Spatial structure of the human alpha rhythm: global correlation in adults and local correlation in children. Clin Neurophysiol. 1999 Aug;110(8):1351–62.

Srinivasan R, Nunez PL, Silberstein RB. Spatial filtering and neocortical dynamics: estimates of EEG coherence. IEEE Trans Biomed Eng. 1998 Jul;45(7):814–26.

Thatcher RW.Cyclic cortical reorganization during early childhood. Brain Cogn. 1992 Sep;20(1):24–50.

Thatcher RW, Krause PJ, Hrybyk M. Cortico-cortical associations and EEG coherence: a two-compartmental model. Electroencephalogr Clin Neurophysiol. 1986 Aug;64(2):123–43.

Thatcher RW, Walker RA, Giudice S. Human cerebral hemispheres develop at different rates and ages. Science. 1987 May 29;236(4805):1110–3.

van Baal GC, Boomsma DI, de Geus EJ. Longitudinal genetic analysis of EEG coherence in young twins. Behav Genet. 2001 Nov;31(6):637–51.

Yakovlev PI, Lecours AR, in Regional Development of the Brain in Early Life, A. Minkowski, (Ed.), Blackwell Scientific, Oxford, pp. 3–70, 1967.

Brain Connectivity and the Spread of Epileptic Seizures

John G Milton[1], Sozari A Chkhenkeli[2] and Vernon L Towle[2]

[1] Joint Sciences Department, The Claremont Colleges, Claremont, CA 91711
jmilton@jsd.claremont.edu
[2] Department of Neurology, The University of Chicago

The occurrence of an epileptic seizure is the result of a paroxysmal and uncontrolled discharge of an enormously large population of neurons. Here we focus on those seizures, referred to by clinicians as secondarily generalized seizures[3], that initiate in a spatially localized epileptic focus and subsequently spread to involve a large portion of the cortex. As the spatial extent of the neuronal population involved in the seizure increases, it becomes natural to question how activity spreads from one point to the other. The apparent simplicity of this question is deceptive. Indeed, obtaining the answer(s) requires a major change in the way that we think about epilepsy. On the one hand, the attention of basic scientists is shifted away from the dynamics of individual neurons towards considerations of how large systems of neurons interact. On the other hand, clinicians must begin to think in terms of an epileptic system (Chkhenkeli and Milton 2002), i.e. a dynamic, ever-changing and evolving, distributed system of neural aggregates that control the onset, propagation, and arrest of epileptic seizures.

The words "spread", "propagation", and "synchronization" of epileptic seizure and epileptic activity have been used for more than a century to facilitate communication between health professionals and basic scientists. The invention and development of the electro–encephalogram (EEG) reinforced this termonology, in part, because the EEG changes during a seizure seemed visually to capture what the terminology attempted to describe. However,

[3] The term "epilepsy" is used when seizures have the tendency to recur. To the non–clinician the terminology used by neurologists to describe seizures can be quite confusing. Here is a simplified overview. The classification of seizures is based on how the seizure begins in the first split second. If the seizure begins in a focal region of the brain it is called partial, if it begins generally everywhere at the same time it is called generalized. Partial seizures are further sub-divided into those in which the level of consciousness is not impaired (partial simple seizures) and those in which it is (partial complex). The term "secondarily generalized seizure" refers to a generalized seizure that was initiated in a spatially localized epileptic focus and then spread to involve a significant portion of the cortex.

recent data suggest that phenomena that occur in association with seizures are actually not well described by these terms. This is one of the main points of this chapter. Nonetheless, during the chapter we will continue to use these habitual words, bearing in mind that they do not fully embrace the complexity of the phenomena that they purport to describe.

The identification of the pathways for the spread of partial seizures is of immense relevance to the subject of human brain connectivity. Not too many years ago, neuroscientists thought that any two neurons in the central nervous system could be connected by pathways that contained no more than three synapses (Guyton 1976). Given that there are $\sim 10^{10} - 10^{11}$ neurons and $\sim 10^{13} - 10^{14}$ synapses (Stevens 1965) in the central nervous system, this would imply that there must be a huge number of pathways by which seizures could secondarily generalize from an epileptic focus. The study of the spread of seizure activity provides a unique opportunity to identify those pathways that have the major functional significance. Although abnormalities in circuitry do arise within the epileptic focus (Engel 1990; Sutula et al.1989) (e.g. sprouting of axon branches and terminals, pruning of dendritic branches and spines), the pathways by which a seizure spreads are often normal. Thus we can anticipate that the pathways for seizure propagation from an epileptic focus will be the same pathways used for information flow within the nervous system under normal circumstances.

Clinical investigations related to seizure onset and spread are typically performed in the context of the pre–surgical evaluation of patients with medically intractable epilepsy. Our goal is to present selected examples obtained during clinical investigations in order to motivate a discussion between clinicians interested in epilepsy, basic scientists interested in brain connectivity, and computational neuroscientists interested in the effects of connectivity on the spatio–temporal dynamics exhibited by large–scale populations of neurons.

1 EEG Signatures of Epilepsy

The primary tools available to monitor the spread of a seizure in a patient with epilepsy are clinical observation and the EEG. The EEG represents a spatiotemporal record of the variations in potential (measured in microvolts) recorded from the brain (Ferree and Nunez 2006; Nunez 1981; Nunez 1995; Nunez and Srinivasan 2006). A rich variety of EEG changes are seen in patients with epilepsy (Fig. 1). The morphology, frequency and spatial distribution of these changes are used by epileptologists to diagnose the epilepsy. Focal epileptic spikes (Fig. 1a) are frequently recorded from patients with partial epilepsy and provide an important clue for the possible location of the epileptic focus. On the other hand, patients with generalized epilepsy exhibit a variety of characteristic generalized EEG waveforms that have diagnostic significance: 3–Hz spike–and–wave patterns in patients with absence seizures (Fig. 1b), the lower frequency spike and wave patterns seen in patients with

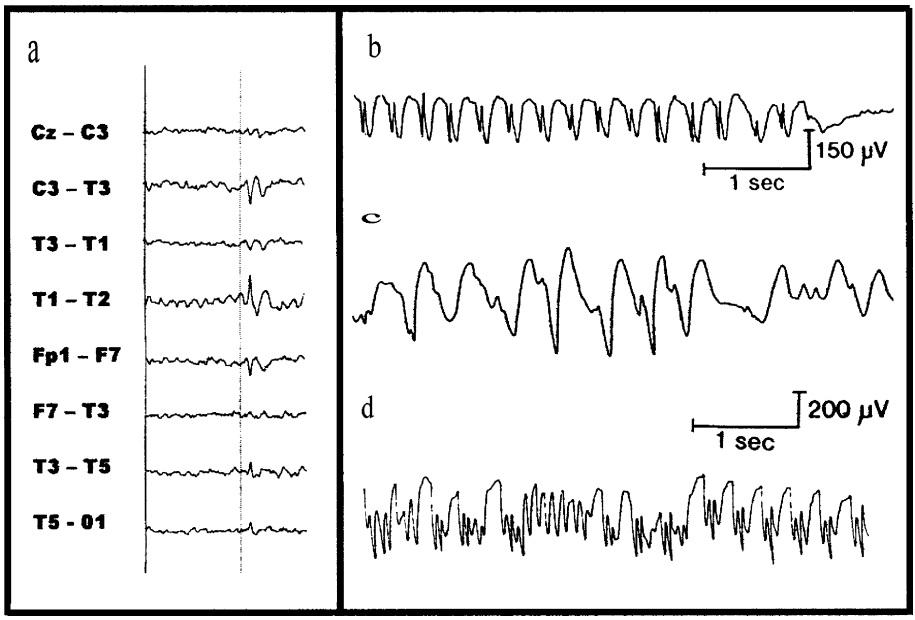


Fig. 1. Examples of EEG changes recorded from patients with epilepsy. (**a**) Isolated epileptic spike recorded from left temporal lobe in a patient with partial complex seizures; (**b**) train of 3 Hz spike–and–wave discharges recorded from a patient with absence seizures; (**c**) train of 2–2.5 Hz slow spike and slow discharges recorded from a patient with Lennox Gastaut syndrome; (**d**) polyspike spike and slow wave activity recorded from a patient in absence status epilepticus

Lennox–Gastaut syndrome (Fig. 1c), and the polyspike and slow waves complexes seen in patients with absence status epilepticus (Fig. 1d). Despite considerable progress in our understanding of the basic molecular and cellular mechanisms of the epilepsies it has not yet been possible to translate this information into a deeper understanding of the EEG phenomena measured on a daily basis in EEG laboratories and epilepsy monitoring units throughout the world.

Little attention has been given to the evolution of EEG patterns, i.e. changes in frequency and morphology of waveforms, as a seizures progresses. One example, that has been emphasized in the clinical literature (Ebersole and Pacia 1996; Pacia and Ebersole 1997), arises in the setting of partial complex seizures in adults that originate from the amygdala–hippocampal–temporal lobe complex. Three different ictal patterns have been described using scalp EEG: a regular and well modulated 5–6 Hz pattern, a pattern characterized by slower rhythms (2–5 Hz) and patterns characterized by diffuse slowing or arrhythmic patterns. These patterns are not unique to the site of seizure onset, but can be observed, for example, in consecutive seizures from the same patient (Chkhenkeli and Milton 2002). The mechanism that produces

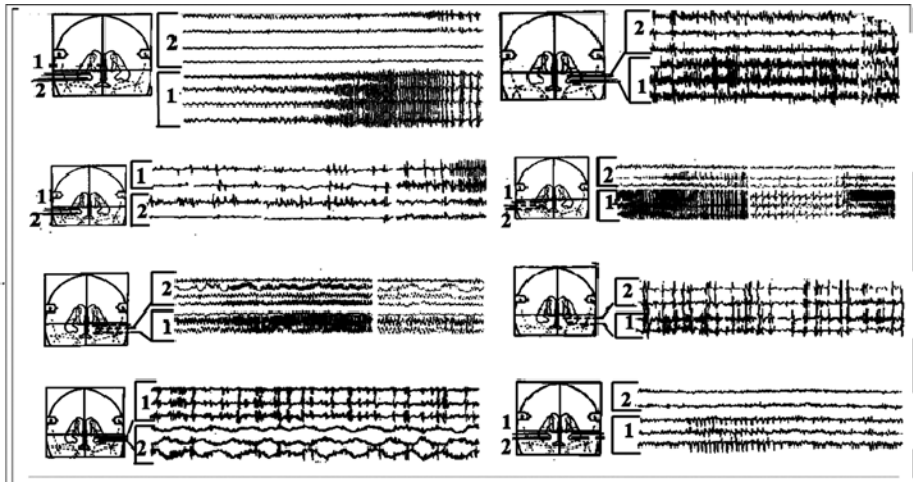


Fig. 2. Different variants of ictal hippocampal activity recorded in eight patients with partial complex seizures. The depth electrodes were implanted in the hippocampus using stereotaxic procedures. '1' refers to channels that record from the hippocamous; '2' to recordings from the amygdala

these ictal EEG patterns is not well understood. Figure 2 shows the EEG recorded using depth electrodes implanted in the hippocampus at seizure onset recorded in eight different patients. Clearly, great variability is observed in the EEG at seizure onset. Thus it is quite likely that the different ictal EEG patterns measured using scalp electrodes arise from the subsequent spatiotemporal dynamics and development of synchrony in neuronal populations as seizure activity spreads from the epileptic focus. Indeed, the observation of a faster and a slower rhythm is quite reminiscent of the fast ($\sim$ 9 Hz) and slow ($\sim$ 3 Hz) bulk oscillations observed in computer simulations of randomly connected networks of integrate–and–fire neurons (Farley 1965). The fact that the oscillations in the EEG and the bulk oscillations in the randomly connected neural networks are considerably slower than the spiking frequencies of typical cortical neurons, e.g. 10–100 Hz, suggest that these EEG patterns are related to emergent statistical properties of large populations of neurons. However, it is not known whether these statistical properties are generated entirely within the cortical mantel or also involve interactions between cortical and subcortical pacemakers.

Since depth electrodes offer the best opportunity for the precise localization of an epileptic focus, the curious reader might wonder why these procedures are not more widely used. The problem with depth electrodes is that they record only from small localized areas of tissue adjacent to the tip of the electrode. Thus, it would be necessary to implant an enormous number of depth electrodes in order to cover all of the brain structures responsible for the onset of a seizures. The current interest in deep brain stimluation for

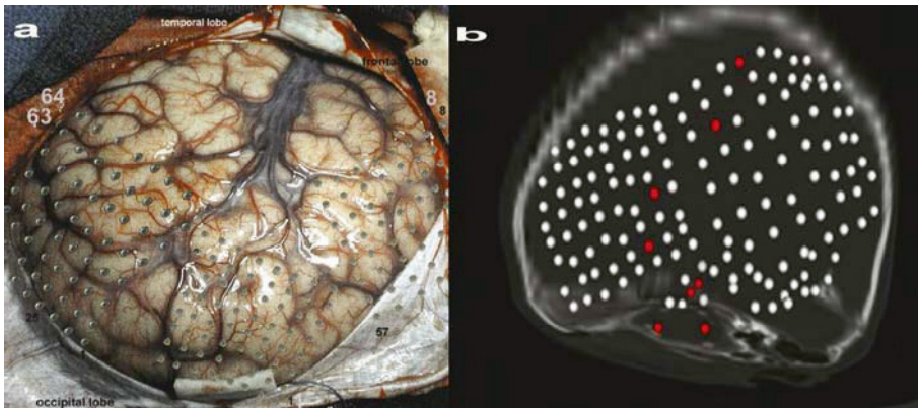


Fig. 3. (**a**) Photograph of the exposed surface of brain revealing the placement of three 8×8 grids of electrodes. Electrode 64 is over the left inferior temporal lobe; the prominent blood vessels indicate the left Sylvian fissure. Pertinent clinical history is given in the legend to Fig. 9. (**b**) CT scan showing the placements of subdural grids of electrodes. ECoG recordings using these grids are shown in Fig. 8. Pertinent clinical history is given in the legend to Fig. 8

the treatment of seizures and other neurological diseases such as Parkinson's disease has produced a resurgence in the use of depth recordings (Chkhenkeli and Chkhenkeli, 1997; Lüders 2004; Milton and Jung 2002).

A practical compromise between scalp and depth electrodes involves the surgical implantation of grids of electrodes (inter–electrode spacing ranging from 2 mm to 10 mm) in the subdural space over the surface of the brain (Fig. 3a). The term 'electro–corticography (ECoG)' is used to refer to EEG recorded directly from the surface of the brain. The location of these electrodes relative to the skull and cortical surface can be determined post-implantation using radiological techniques (Fig. 3b). An advantage of subdural over scalp EEG recordings for the study of seizure propagation is that such recordings are relatively free from motion and scalp muscle artifact: the grids can be attached with a stitch to the dura mater.

The combination of ECoG with modern techniques for neuro–imaging and time series analysis makes it possible to visualize the spread of seizures on the scale of a signification portion of the brain. Figure 4 shows a seizure recorded in a patient with multifocal onset partial epilepsy in whom four subdural grids of electrode were implanted over the left hemisphere: an anterior grid implanted over the frontal lobe, a posterior grid over the parietal–occipital lobe, and two strips of electrodes inferiorly under the temporal lobe. As the seizure spreads, the frequency and morphology of the EEG recorded from the electrode nearest the presumed epileptic focus (colored red) continually changes. The green lines in the figure indicate the electrode pairs for which the EEG is significantly coherent. The important observation for this discussion is that the EEG activity between two electrodes located a great distance apart

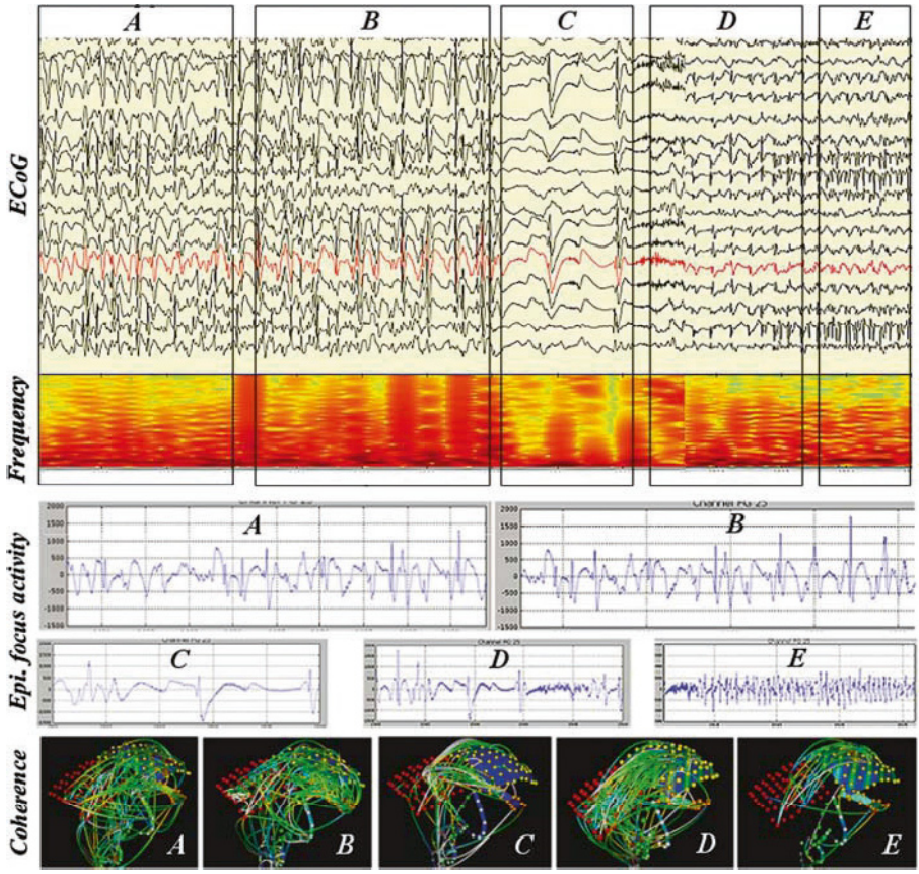


Fig. 4. ECoG recorded during a seizure onset from a patient with multifocal seizure onset using four implanted grids of subdural electrodes: the frontal grid electrodes are colored red, the parietal–occipital grid electrodes are yellow, and the two temporal electrode strips are green. The top row of panels show ECoG recorded from 17 of the total of 128 channels. The ECoG is 36 second long and has been subdivided into five sections: **A**: interictal epileptic activity; **B**: increased spike activity; **C**: electrodecremental stage, **D**: seizure onset, **E**: later stage of seizure. The second row of panels show the changes in the ECoG power recorded from the electrode cloest to the presumed epileptic focus for the shown seizure (red). The x–axis of each panel shows the time measured in seconds, the y–axis shows the frequency content (0 Hz at the bottom, 100 Hz at the top, and a color scale is used to represent the ECoG power at each frequency (red is high power, blue is low). The third row of panels show examples of the ECoG recorded from the electrode placed closet to the epileptic focus (the red ECoG tracing shown in the top row of panels). The bottom row of panels summarizes the coherence between measured between two electrode pairs: a green line connects those electrode pairs which had significant coherence (Fuchs 2007; Kamiński 2007). The voltages were recorded relative to the contralateral mastoid. The grand mean was subtracted from each electrode before the coherence was calculated

(the electrodes on the grid are spaced at 1 cm intervals) can be coherent even though coherence is not seen for the intervening electrode pairs.

1.1 Generation of the EEG

In order to better understand the significance of the EEG changes associated with a seizure it is useful to have an understanding of how the EEG is generated. The EGG measured using scalp or subdural electrodes is a measure of the activity of neurons which are located in a thin circumferential cortical band measuring 1.5–4.5 mm in thickness that is located directly at the surface of the brain The neurons are not scattered in a haphazard manner but are arranged according to two major organizational features: horizontal lamination and vertical columnation. To a first approximation, the EEG is generated by the large vertically oriented pyramidal neurons located in layers II, III, V and VI. These neurons have two features that are critically important for the EEG: 1) they are oriented perpedicular to the cortical surface; and 2) the inhibitory and excitatory inputs are spatially segregated over the surface of these neurons. In particular, only inhibitory synapses are made at the soma. On the other hand, dendrites receive both excitatory and inhibitory synapses: most of the synapes made onto dendritic shafts (axodendritic) are inhibitory, whereas those onto spines (axospinous) are usually excitatory.

The generators of the EEG are the excitatory and inhibitory postsynaptic potentials, respectively the EPSPs and the IPSPs, that summate on the pyramidal neurons (Fig. 5). Consider the effect of increasing excitation on the more distal part of the apical dendritic tree. By definition, the direction of current flow is taken with respect to the direction that the positive ions flow. Since an EPSP causes an inward flux of positive ions, the extracellular compartment at this point in space becomes more negative. To maintain neutrality this inward flux of positive ions must be balanced by an outward flux of positive ions at the soma. Hence, the deeper extracellular compartment becomes more positive. Consequently, the pyramidal neuron can be thought of as a vertical dipole oriented so that the sink (negative pole) is near the cortical surface and the source (positive pole) located slightly deeper from the surface (i.e. $\leq 1{-}2$ mm) (Darvas and Leahy 2007). If, on the other hand, excitatory input is increased to synapses located closer to the soma, the dipole has the reverse orientation. By repeating these arguments it is easy to see that the effects of increasing IPSPs at deeper cortical layers is identical to increasing EPSPs at more superficial layers: increasing IPSPs at superficial levels is the same as increasing EPSPs at deeper levels.

The potential generated by a single neuronal dipole is too small to be measured by an EEG electrode. The reason why these very small extracellular currents can be measured is because the pyramidal neurons are aligned in a parallel fashion throughout the gray matter. This anatomical feature means that individual dipoles can summate to produce an equivalent dipole large enough to be measured. Thus, the EEG is generated by a dipole layer of neurons.

There are three important consequences of this dipole layer theory for the generation of the EEG. First, the potential generated by a dipole layer falls

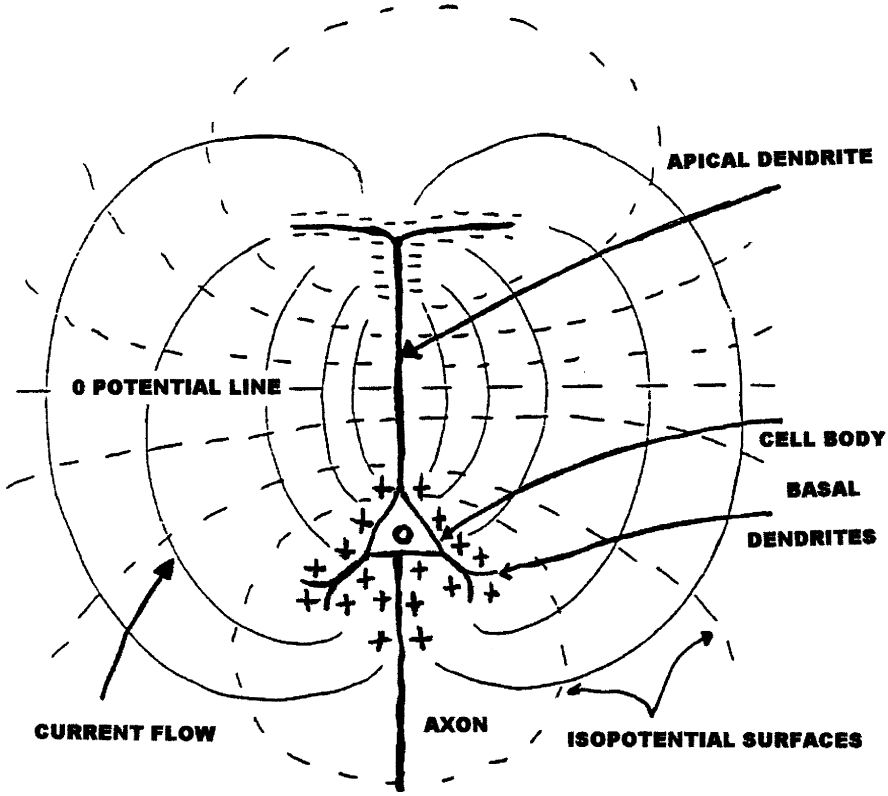


Fig. 5. Schematic representation of a pyramidal neuron. Reproduced from (Ebersole and Milton 2002) with permission

off less quickly with distance than that generated by a single dipole (Lopes da Silva and van Rotterdam 1982). Second, the amplitude of the EEG recorded from a cortical region in which neurons are not parallel will be much lower than that recorded from cortical regions having vertically oriented neurons. This is because the dipoles generated by a population of randomly oriented dipole will tend to cancel. Since neurons in subcortical neurons tend not to be aligned in parallel, this means that the activity of these neurons do not meaningfully contribute to EEG recorded by electrodes placed on the cortical surface. Finally, the EEG is essentially a measure of synchrony of the neuron dipoles within the dipole layer. The time–averaged potential, $|\overline{\Phi}|$, for m parallel dipoles which oscillate in phase is (Nunez 1981)

$$|\overline{\Phi}| \sim m$$

whereas if they oscillate with random phase it is

$$|\overline{\Phi}| \sim \sqrt{m}$$

Since for a scalp electrode there are no local sources, the relative contribution of the m synchronous and n asynchronous oscillating dipoles to $|\overline{\Phi}|$ is $m/\sqrt{n}$ (Nunez 1981), the EEG emphasizes the contribution of the synchronously oscillating dipoles (Ebersole and Milton 2002).

1.2 Synchronization, or something else?

A word that is frequently abused in discussions of neural activity during a seizure is *synchronization*. Historically it was felt that a seizure represented a state of "hypersynchrony", i.e. synchronous activity in a very large population of neurons (Penfield and Jasper 1954). However, it has proven to be very difficult to directly demonstrate synchrony between the firing of individual neurons during a seizure (Babb et al. 1987; Netoff and Schiff 2002; Wyler et al. 1982). For example, in animal models with penicillin–induced generalized 3 Hz spike–and–slow wave epilepsy, the neurons do not typically fire synchronously (Steriade and Amzcia 1994; Steriade and Contreras 1998). Even during ongoing human partial complex seizures, depth electrode recordings from the ipsilateral hippocampus indicate that less than 30 % of the neurons change their firing frequency (Babb et al. 1987). Indeed recent studies of epileptic seizures occuring in hippocampal slices (Netoff and Schiff 2002) and computer simulations (van Drongelen et al. 2005) have argued that desynchronization of neurons may be more relevant to the occurence of an epileptic seizure.

There are three issues that have contributed to this confusion. First, epileptologists do not usually recognize the fact that synchrony is not simply limited to the situation in which all neurons spike at the same time. Synchronization describes the adjustment in rhythms that occur as a result of interactions between oscillators (Pikovksy et al. 2003). From this perspective, complete coincidence of signals is but one of a rich array of possibilities that range from a variety of phase locking patterns (Bressloff and Coombes 2002), some of which are very complex, to phase synchronization, i.e. populations of coupled chaotic oscillators that are synchronous in phase but not in amplitude (Rosenblum et al. 1996). Although new techniques to detect synchronization are increasingly being used to study epileptic seizures, sometimes with great effect (Dominguez et al. 2005), it is important to realize that these approaches are not without their own problems (So et al. 2002).

Second, scientists have tended to overlook the fact that epileptologists have traditionally used the word synchronization in the context of the dipole layer mechanism that generates the EEG. As we have seen this dipole layer theory does not necessarily imply that all neuronal dipoles oscillate in phase, only that a certain fraction of them do. Consequently, it can be very difficult to relate rhythmic changes recorded in the scalp EEG with changes in the spiking patterns of the underlying neural populations. An important exception is the high correlation that exists between the probability of a spike occuring

in a single neuron and the waveform and amplitude on an evoked potential measured using EEG (Fox and Norman 1968).

Finally, and perhaps more importantly, both scientists and epileptologists often overlook the fact that the EEG is a measurement of the statistical properties of a large population of neurons (Milton 1996; Milton and Mackey 2000). It has been estimated that a typical scalp EEG electrode records from about 5–10 cm^2 of the underlying cortical surface, i.e. approximately $\sim 10^8$ neurons (Cooper et al. 1965; Tao et al. 2005). From this perspective the potential measured by a single EEG electrode corresponds to the average activity of the underlying population of neurons: rhythmic changes in the EEG potential imply that cycle changes occur in the average activity. Theoretical (Lasota and Mackey 1994; Milton and Mackey 2000) and experimental (Cole 1991) investigations have shown that it is possible for the population average to be cyclic even though no element in the population itself exhibits periodic activity! One possible explanation is *statistical periodicity*, i.e. the probability density function cycles in a periodic manner (Lasota and Mackey 1994; Milton and Mackey 2000). This phenomenon occurs only in time–delayed dynamical systems. Hence one would be anticipate that statistical periodicity might be very relevant for populations of neurons since axonal conduction velocities are finite. Statistical periodicity is a very robust phenomenon: it arises in populations of chaotic elements and in populations of linear dynamical elements subjected to the effects of either state-independent (additive) or state-dependent (multiplicative) noise (Lasota and Mackey 1994). However, the relationships between statistical periodicity, synchronization, and the EEG, if any, have not yet been established.

2 Seizure onset

Patients with epilepsy always ask their physician, "Why do seizures occur?" Perhaps the more astute might ask why doesn't everyone have epilepsy! Indeed, from an anatomical point of view the cortex appears to be set up as a time bomb just waiting to go off. Approximately 75 % of the cortical neurons are excitatory. Each neuron receives $\sim 10^3 - 10^4$ inputs (Stevens 1965) and the excitatory inputs outnumber the inhibitory ones by 6.5 : 1. (Beaulieu et al. 1992). Local excitatory input arises from the collaterals of the excitatory neurons themselves and ramifies through the laminae as well as both upward and downward to create a rich arborization of local excitatory input. Thus it is not surprising that mathematical studies (Traub and Miles 1991; Traub et al. 1989; Wright 1999) and experimental observations (Buzsáki 1986; Buzsáki 1989; Buzsáki et al. 1987) have indicated that the isolated cortex is operating near a threshold that separates quiescent behavior on the one hand from an active epileptic state on the other hand. These observations suggest that the inhibitory mechanisms present in the cortical layer of healthy brain are not sufficient by themselves to maintain the cortex in a non-seizing state.

The episodic nature of seizure occurence (Milton et al.1987) implies that the epileptic focus may face a similar problem for maintaining an ever teetering balance made even more difficult because of the pathological alterations to its control mechanisms.

It has long been known that the onset of partial seizures can be heralded by a flattening of the EEG (see, for example, section C in Fig. 4). Historically this EEG flattening, or electrodecremental response, was interpreted in terms of a reorganization of neuronal dipoles. With modern EEG recording technologies it has been shown that high frequency (100–500 Hz) oscillations occur during this period, particularly in electrodes placed in the region of the suspected epileptic focus (Bragin et al. l999; Jirsch et al. 2006; Schiff et al. 2000; Worrell et al. 2004). Initially it was thought that the spatial extent of these high frequency oscillations was small, i.e. < 1 mm (Bragin et al. l999; Bragin et al. 2002); however, it is now known that they can be detected using electrodes of various sizes (Jirsch et al. 2006; Worrell et al. 2004). This observation implies that high frequency oscillations may exist on a variety of length scales perhaps up to the order of the size of the epileptic focus but certainly less than that of the whole cortex. Gap junction connections between principle neurons may play an important role in the generation of these high frequency oscilations (Bragin et al. 2002; Traub et al. 2001).

Brief sensory (Milton 2000) and electrical (Lesser et al. 1999; Motamedi et al. 2002) stimuli applied near seizure onset can abort the seizure. These observations provide very suggestive evidence for multistability (Beggs et al. 2007; Campbell 2007; Milton 2000), i.e. the coexistence of a quiescence and a seizing state of the cortex. The possibility of multistability was first suggested by computational studies of studies of model populations of neurons which indicated that the onset of traveling waves was characterized by a subcritical bifurcation (Chu et al. 1994; Milton 1996; Milton et al. 1993; Milton et al. 1995), i.e. by a bifurcation associated with the coexistence of two or more qualitatively different activity states. An understanding of how seizures initiate may led to the development of new treatment strategies for epilepsy (Milton and Jung 2002).

A more modern concept has been to explore the implications of a cortex stabilized near, or at, "the edge of stability" by subcortical inputs (Buzsáki and Traub 1997; Milton and Jung 2002; Milton et al. 2004). Although electrical stimulation of the sub–cortical nuclei does not result in the generation of epileptic seizures (Ajmone–Marsan 1972), both animal studies and studies in humans with depth electrodes have demonstrated that three structures: thalamus, cingulum, and midbrain, are so important that seizure spread is not possible without their participation (Chkhenkeli and Milton 2002). The advantage of tuning a control system near the edge of stability is that it can respond to perturbations very quickly (Cabrera and Milton 2002; Cabrera et al. 2004; Chialvo and Bak 1999; Kelso 1995; Langdon 1990; Moreau and Sontag 2003). Presumably this advantage outweighs the risk of the occurence of a seizure. An attractive aspect of the hypthesis that cortical–subcortical

interactions are important for epilepsy is that it readily explains why any person, under the appropriate circumstances, can experience a seizure and why so many subcortical structures exist in the healthy nervous system that are capable of inhibiting seizure generation and generalization (Chkhenkeli 2002; Chkhenkeli and Milton 2002; Chkhenkeli et al. 2004; Lüders 2004), e.g. cerebellum, caudate nucleus, centromedian nucleus of the thalamus, the raphe nuclei, the locus coeruleus, substantia nigra and the nucleus of the solitary tract.

If the cortex is indeed tuned at the edge of stability then one would anticipate that cortical dynamics, including seizure onset, should demonstrate the same dynamical phenomena observed for other dynamical systems tuned near a stability boundary (Milton et al. 2004; Worrell et al. 2002). These properties, collectively referred to as *critical phenomena*, include the appearance of power laws, intermittency, multistability, and critical slowing down. Current interest has focussed on determining whether long-range correlations with power-law scaling occur in cortex. Such correlations are expected findings for relaxational processes, i.e. processes which occur whenever a critical variable exceeds a certain critical threshold. It has been shown that fluctuations in EEG energy in both normal (Linkenkaer–Hansen et al. 2001) and epileptogenic (Cerf et al. 2004; Parish et al. 2004; Worrell et al. 2002) brain exhibit long range temporal correlations with power law scaling. Theoretical arguments have suggested that neural systems at the edge of stability are expected to exhibit a power law with exponent $-3/2$ (Cabrera and Milton 2002; Chialvo and Bak 1999; Eurich et al. 2002; Milton et al. 2004). Although such power laws have been observed in a variety of neural experimental paradigms including bursting behavior in cultured neurons (Segev et al. 2002) and brain slices (Beggs and Plenz 2003) and in motor tasks controlled by the intact nervous system (Cabrera and Milton 2002; Cabrera et al. 2004), they have not yet been reported for cortical activity in the intact brain.

In summary, the presence of long range correlations in EEG data with power law exponents suggests that seizure onset may share features exhibited by physical systems such as the water–ice phase transition. Computer simulations of water freezing indicate that the nucleation sites for ice formation initially expand slowly until they reach a certain critical size after which they expands on a much faster time scale (Matsumoto et al. 2002) What corresponds to the nucleation sites for a seizure? Perhaps the localized high frequency oscillations observed at seizure onset will provide a clue.

3 Seizure Spread in Slices of Disinhibited Brain

Propagation of activity in spatially extended populations of randomly interconnected excitable cells is expected to take the form of waves of neural activity (for introductory reviews see (Milton 1996; Milton et al. 1995). In this situation, the nature of the waves, i.e. standing or traveling, depends primar-

ily on the relative spatial extent of the inhibitory and excitatory connections. When the spatial range of inhibition exceeds that of excitation, standing wave patterns predominate. On the other hand, traveling waves, e.g. target waves and spirals, occur when the spatial range of excitation is relatively larger (Fig. 6a). These patterns depend on the strength of the interneuronal excitation. Although it is difficult to alter the spatial extent of the excitatory connections in neocortex experimentally, the spatial extent of the inhibitory connections can be readily changed. To understand why this is true it is necessary to remember that the integrate–and–fire properties of neurons imply that there are two distinct mechanisms that produce a decrease in the rate of neural spiking. First, the resting membrane potential can be lowered (hyperpolarized). This effect is typically the result of synaptic transmission involving inhibitory interneurons, and hence can be blocked pharmacologically using, for example, bicuculline, carbachol and/or penicillin. Second, the threshold for spiking can be increased to produce the absolute and relative refractory periods. Refractoriness occurs intrinsically in all spiking neurons to some extent following the generation of a single action potential. Thus, when the neocortex is disinhibited pharmacologically the spatial range of the excitatory connnections necessarily exceeds that of the inhibitory ones and consequently traveling waves are expected to occur.

Initial attempts to demonstrate the presence of spiral traveling waves in disinhibited slices of mammalian cortex (Chen et al. 2000; Friedrich et al. 1991; Fuchs et al. 1987; Iijima et al. 1996; Petsche et al. 1974) and intact

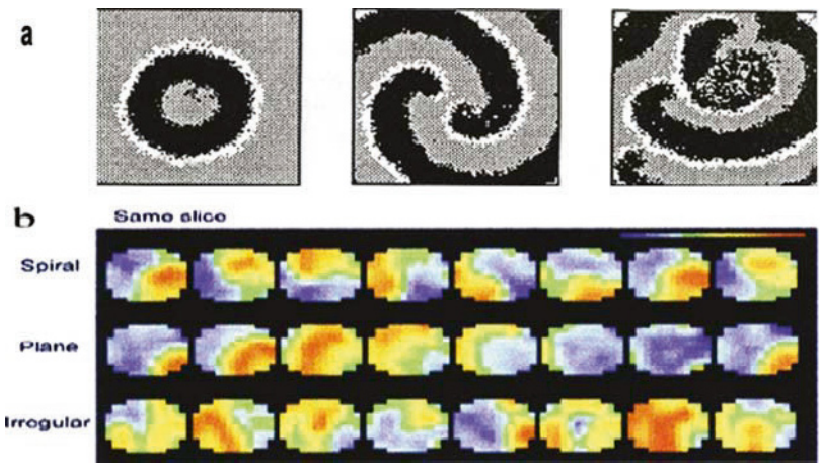


Fig. 6. (**a**) Spatio–temporal activity generated in excitatory networks of integrate–and–fire neurons (Chu et al.1994; Milton et al.1993). (**b**) Wave patterns measures in tangential slices from rat occiptal cortex. Spiral, plane, and irregular waves are seen from the same field of one slice. Colors represent the amplitude of the signal following a linear color scale (top right). Images are 1 msec snapshots of the waveforms. Reproduced from (Huang et al. 2004) with permission

human cortex (Haglund 1997; Haglund et al. 1992) were disappointing. Recently, Huang and co–workers (Huang et al. 2004) successfully demonstrated the presence of traveling spiral waves in disinhibited slices of rat occipital cortex using voltage sensitive dyes (Fig. 6b). The wavelength of the observed spiral waves is of the order of $\leq 1-2$ mm; in excellent agreement with that predicted from computer simulations of networks of integrate–and–fire neurons, i.e. $\leq 1-4$ mm (Chu et al. 1994; Milton et al. 1993; Milton et al. 1995). A key step was that the neocortical slices were tangentially sectioned so that the sectional plane was parallel to the cortical laminae, and hence preserved the horizontal connections in layers III–V. It is quite possible that the failure of previous investigators to observe spiral waves in mammalian cortex was because the tissue used was not sufficiently smooth and isotropic as in heart (Winfree 2002).

The above observations indicate that the spatio–temporal dynamics of disinhibited spatially extended networks of neurons closely resemble those that arise in other excitable media (Winfree 2002). However, inhibitory mechanisms are always active in intact cortex and even in the epileptic focus (Babb et al. 1989; Haas et al. 1996). Traveling wave patterns similar to those shown in Fig. 6 have not been demonstrated in cortical slices when inhibition is present (Chervin et al. 1988; Rinzel et al. 1998).

A related question concerns the nature of the wave–like properties of neural activity that would be detected by the EEG (Nunez 1981; Nunez 1995; Nunez and Srinivasan 2006). Provided that the neocortex has a closed topology, i.e. all cortical regions are interconnected, it been possible to show that all weakly damped traveling waves will tend to develop into standing waves on appropriate scales of length and time (Nunez and Srinivasan 2006). An important conclusion of this analyses is the demonstration that the properties of EEG waves can depend on how the EEG was measured. In particular, the global nature of these calculations implies that they are most valid the larger the neural population monitored by the EEG electrode (Nunez and Srinivasan 2006). In other words, different conclusions can be reached depending on whether, for example, the EEG was measured using scalp, subdural or depth electrodes.

4 Seizure Spread in Intact Brain

For many the term 'seizure spread' conjures up the mental picture of abnormal neural activity spreading out from an epileptic focus progressively engulfing more and more neurons "like a diffusing oil drop" (Kreindler 1965). Unfortunately, the image of a diffusing oil drop is not always a totally valid representation for the spread of seizures in the intact human brain. In fact, the study of seizure spread provides a most convincing demonstration of the importance of getting the brain connectivity right before speculating on mechanism! There are four features of the EEG recorded during secondarily generalized seizures

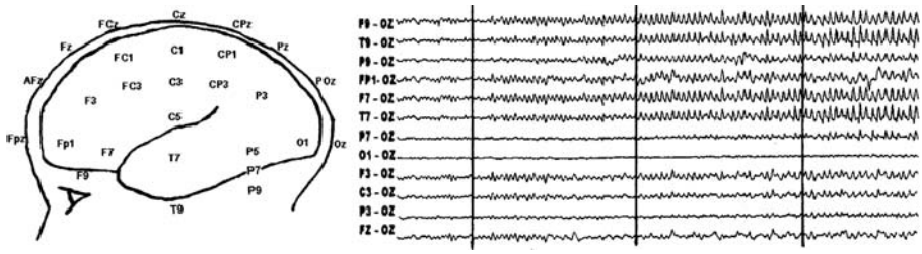


Fig. 7. Scalp EEG recording of a partial complex seizure arising from the left amygdala–hippocampal complex. The electrode placements are shown in the schematic on the left–hand side of the figure. Vertical bars delineate a time interval of 5s. The EEG recording is adapted from (Ebersole and Milton 2002)

that provide clues for understanding how seizures spread in the intact human brain.

First, during seizure propagation the expected lag between signals recorded near the epileptic focus and far from it is much smaller than would be anticipated from estimates of the intracortical (horizontal) seizure propagation velocity of $\leq 6-18$ cm/sec. For example, consider the partial complex seizure recorded in Fig. 7. This seizure arises in the amygdala–hippopcampal complex spreads to the surface of the temporal lobe via the temporal pole (Buser and Bancaud 1983; Buser et al. 1972). In terms of electrodes (left side of Fig. 7), the seizure would propagate, for example, from $F_7 \rightarrow T_7 \rightarrow P_9$. These electrode placements are approximately 2-3 cm apart, and hence we would expect the rhythmic activity recorded at P_9 to lag that recorded at F_7 by $\sim 280-850$ m sec. However, as can be seen, the actual lag is much smaller, i.e. approximately zero. Indeed, even when EEG changes are compared between the two hemispheres during a secondarily generalized seizure, the lag is typically ≤ 20 msec (Gotman 1983). These observations clearly imply that there must be a faster route for seizure generalization than intracortical propagation via the horizontally–directed unmyelinated axons of the cortex.

Second, secondarily generalized seizures terminate in a manner that suggests that the epileptic focus is not the primary pacemaker at the time of seizure cessation. Figure 8 shows the termination of a secondarily generalized seizure. Contrary to expectation, the ECoG changes associated with the seizure stop abruptly in the frontal, parietal, and both temporal lobes at approximately the same time. The simplest explanation for these observations is that the secondarily generalized seizure is being driven by a common sub–cortically located pacemaker. This suggestion in turn implies that at some point during the secondary generalization of a seizure, the mechanisms for maintaining the seizure must shift from the cortex (i.e. the location of the epileptic focus) to some other structure, possibly located in the sub–cortex.

Third, seizures can simultaneously involve non–contiguous regions of cortex (see also Fig. 4). Figure 9 shows a ECoG recording of seizure using the subdural grid of electrodes shown in Fig. 3a. For each electrode the Fourier

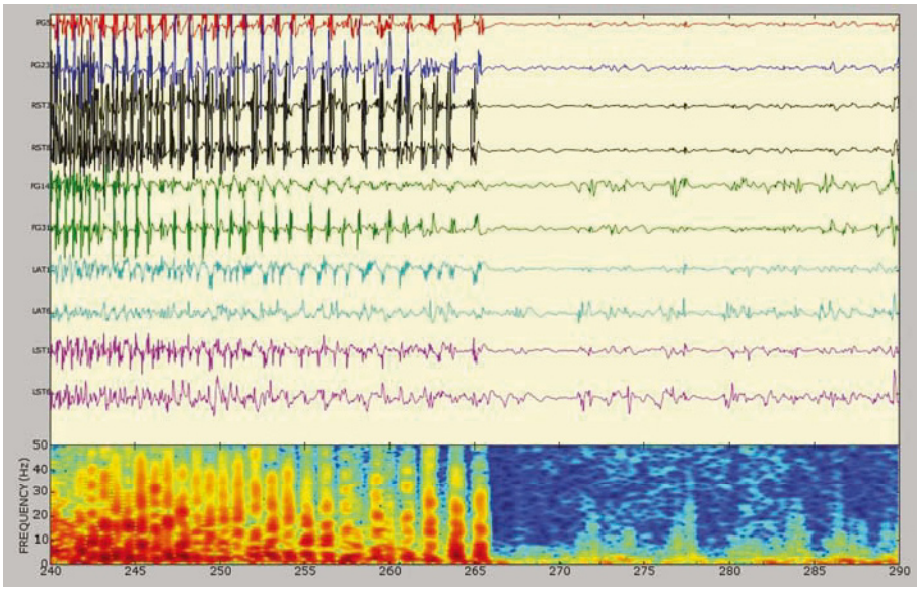


Fig. 8. ECoG recordings of the cessation of a secondarily generalized seizure recorded from a ten–year old female with multifocal seizures (one epileptic focus was located in the left frontal lobe, the other in the left parietal–occipital region. Magnetic resonance imaging (MRI) showed an area of encephalomalacia in the left temporal lobe. The location of the EEG channels correponds to the red electrodes in Fig. 3b. The bottom of the figure shows the ECoG power recorded from the parietal electrode PG5; i.e. the channel at the very top of this figure

power spectrum was calculated and the total power represented in a color scale (red indicates the highest power, blue the lowest). This seizure began in a malformed region of the lateral occipital lobe and then spread to more superior and anterior regions of the posterior temporal lobe. During the seizure there

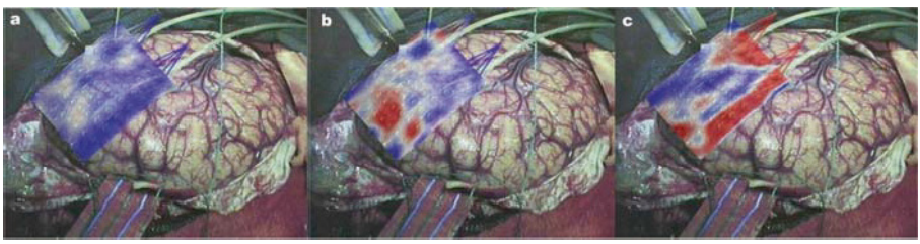


Fig. 9. Changes in ECoG power recorded from the occipital grid (Fig. 3a) before (**a**) and during (**b**,**c**) a seizure recorded from an eight–year old male with left hemimegencephaly involving the left temporal–parietal–occipital region. The seizure (red) begins in the posterior aspect of the grid. A color scale has been used to represent the power: highest (red) to lowest (blue)

were intervening areas in the middle of the grid that did not display epileptic activity. However, the frequency content of the seizing areas was similar suggesting that these non-contiguous areas were nonetheless functioning as a unitary phenomenon. This interpretation was supported by measurements of the coherence between EEG electrodes (Towle et al. 2002).

5 Pathways for Seizure Spread

The observations in Sect. 4 indicate that seizures in the human brain do not significantly propagate via horizontal connections confined to the cortical layer[4]. In order to appreciate the nature of the other pathways available for seizure propagation it is helpful to represent the brain schematically as the double sphere shown in Fig. 10. In this representation the neocortex is represented by the circumference of the outer sphere and the inner sphere encloses all of those sub–cortical and brainstem nuclei that modulate the onset, propagation and subsequent termination of the epileptic seizure. The large area between the circumferences of the inner and outer spheres is the *white matter* and contains all of the cabling, i.e. the myelinated and unmyelinated axons, that interconnect neurons located in the gray matter and sub-cortical regions as well as those that convey the sensory inputs and motor outputs of the central nervous system. It follows that there are three routes by which seizures can spread:

1. Intracortical spread (Path 1 in Fig. 10): epileptic activity spread horizontally from the epileptic focus via thin diameter, unmyelinated axons;
2. White matter mediated spread (Path 2 in Fig. 10): seizure activity propagates from the gray matter vertically via myelinated axons that subsequently arc back towards the surface to interconnect adjacent as well as distant areas of cortex in both the ipsilateral and contralateral hemisphere;
3. Sub–cortical nuclei mediated spread (Path 3 in Fig. 10): seizure activity spreads vertically via myelinated axons to the underlying sub–cortical nuclei which subsequently project more diffusely back to the cortex.

It is quite likely that all of the pathways shown in Fig. 10 participate, to varying extents, in the spread of seizure activity from an isolated epileptic focus. The velocity of seizure propagation along each of these pathways is expected to differ markedly: seizure generalization via intracortical spread is slower than by white matter mediated spread which is slower than spread mediated by sub–cortical nuclei (Milton 2002). The faster a clinical seizure spreads, the more likely that the faster propagation pathways are involved. However, it is clear that there is not necessarily a perferred pathway for a seizure to spread from an epileptic focus; consecutive seizures from the same

[4] A possible exception concerns Jacksonian march seizures that are confined to the primary motor cortex.

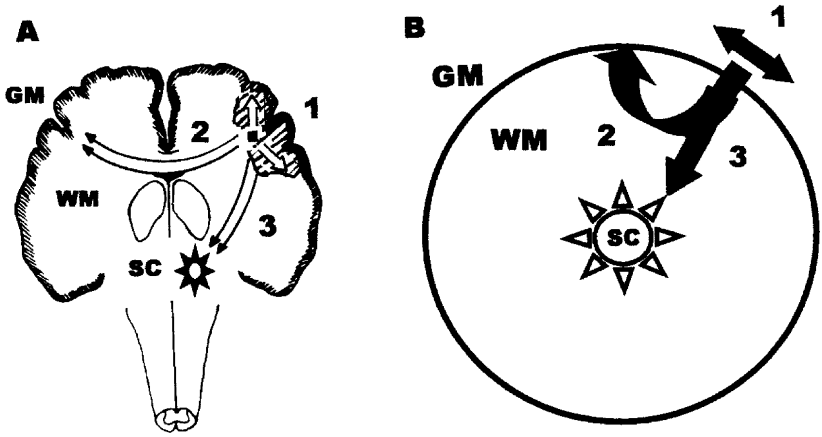


Fig. 10. Schematic representation of three potential seizure propagation pathways (**A**). A simplification of these pathways is shown in (**B**). The starburst symbol for the sub–cortical nuclei (SC) indicates that they have a diffuse projection. WM: white matter, GM: gray matter. Reproduced from (Milton 2002) with permission

patients can spread via different pathways; for an example see (Chkhenkeli and Milton 2002).

An important corollary of the columnar organization of cortex is that the major direction for the flow of neural activity is perpendicular to the cortical surface, i.e. in the direction that the myelinated axons conduct (Petsche et al. 1988). Measurements of spike propagation in human temporal lobes (Baumgartner et al.1995; Emerson et al. 1995; Sutherling and Barth 1989) and between the two hemispheres (Gotman 1983, Ktonas and Mallart 1991), the spread of electrically induced after–discharges in animal models (Kreindler 1965), and of dispersive waves over the cortical surface (Nunez 1981; Nunez 1995) provide convincing evidence that neural activity can indeed propagate via the interconnecting axons that form the association and commissural fibers of the white matter. However, focal electrical stimulation studies in animals demonstrate that cortically generated after–discharges spread preferentially to sub–cortical structures rather than directly to contiguous cortical regions (Ajmone–Marsan 1972; Konigsmark et al. 1958). The lone exception appears to be after–discharges generated in primary sensorimotor cortex which spread first to homologous regions in the contralateral hemisphere.

An illustrative example concerns the spread of seizure activity from one temporal lobe to the other. Intuitively, one might anticipate that the route would be along the anatomical commissures that interconnect the temporal lobes, i.e. the anterior and hippocampal commissures. However, this is not the case (for a review see (Chkhenkeli and Milton 2002)). The major route for seizure propagation between the two temporal lobes is via vertical spread of seizure activity by indirect pathways involving many deep brain structures,

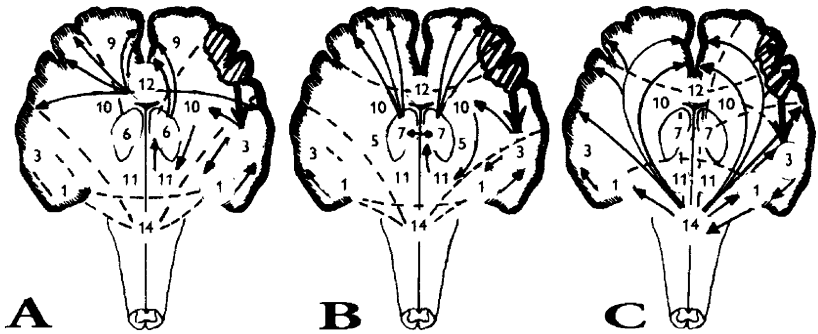


Fig. 11. Diagramatic representation of the main pathways for the spread of epileptic discharge from the right amygdalo–hippocampal complex: (**A**) spread involving the anterior thalamic nuclei, gyrus cinguli and corpus callosum; (**B**) spread involving the non–specific thalamic nuclei; (**C**) spread involving the mesencephlic reticular formation. The *solid lines* show the "main" directions of spread and generalization of the epileptic discharge and the *dotted lines* show the auxillary routes. 1– amygdala, 3–hippocampus, 4–gyrus hippocampus, 5–dorsomedial thalamic nuclei, 6–anterior thalamic nuclei, 7–centrum median thalamic nucleus, 9–gyrus cinguli, 10–fornix, 11–hypothalamus, 12–corpus callosum, 14–mesencephalic reticular formation. Reproduced from (Chkhenkeli and Milton 2002) with permission

including the brainstem. This is a consequence of the fact that the projections of the amygdala–hippocampal complex, the typical site for the epileptic focus in patients with partial complex seizures, are not directed to the symmetric contralateral structures, but rather to ipsilateral deep–brain structures from which the seizure activity subsequently spreads. A variety of possible routes have been discovered; three examples are shown in Fig. 11. The common route (Fig. 11a) is hippocampus → fornix → mamillary body → mamillothalamic tract → posterior cingulate gyrus (Brodman areas 23 and 29) → contraleteral frontal lobe → contraleteral temporal lobe. Another possibility is amygdala → non–specific thalamic nuclei → ipsilateral anterior cingulate gyrus (Brodman area 24) → contraleteral temporal lobe (Fig. 11b). Although often neglected in discussions of seizure spread, structures located in the midbrain likely play important roles (Weir 1964); this is illustrated in Fig. 11c.

6 Discussion

The spread of a seizure is clearly a very complicated phenomenon that involves a dynamic cooperation between neurons located in the cortex with those located in sub–cortical nuclei. However, this cooperation is not restricted to epilepsy: many functions of the brain are actually functions that involve an interplay between cortex and subcortex (Jentzen and Kelso 2007). For neuroscientists focussed on the cortex, it is rather sobering to observe the behavior

of birds (Romer 1967). Despite the fact that a bird brain is dominated by its basal nuclei, birds nonetheless can exhibit a complex series of action patterns appropriately in response to a great variety of situations. The important distinction between birds and animals that possess a relatively larger cortex is that birds have only a very limited capacity to learn. However, it must not be forgotten that the execution of learned action plans formulated by the cortex in animals must take into account the role of the subcortical structures, but also the physics of the system and the task (Chiel and Beer 1997). Thus, it should not be surprising that to learn that although seizures are generated by the cortex (but not always (Kuebler and Tanouye 2000)), their spread is dominated by the cortical–subcortical interactions.

The dynamical landscape over which the seizure evolves is not static, but is rather a continuously undulating one, evolving over times scale ranging from minute–to–minute, to the lifetime of the patient (Milton 2002). The technical problems for studying interactions between cortex and subcortex stem from the fact that these regions are spatially separated: the cortex is closer to the surface of the skull that the subcortex. Although the combination of magnetic resonance imaging with neurophysiological techniques, e.g. EEG or the magneto–encephalogram, is capable of providing high spatial and temporal resolution, their use for monitoring seizure spread is obviously limited by concerns for the safety of a seizing patient while inside the scanner. We believe that it is unlikely, at least in the foreseeable future, that new technologies will become available that can directly monitor the cortical-subcortical interactions that occur during a seizure in a patient with epilepsy.

On the other hand, we believe that a more feasible approach will be to combine existing technologies with computational methods that "fill in the gaps" between what can and cannot be measured (Beggs et al. 2007; Campbell 2007; Breakspear and Jirsa 2007; Horwitz and Husain 2007). For example, the double sphere model for seizure spread shown in Fig. 10 is very reminiscent of paradigms for the feedback control of wave propagation in excitable systems, e.g. (Mihaliuk et al. 2002; Sakurai et al. 2002) and the feedback control of epilepsy (Chkhenkeli 2002; Fountas et al. 2005; Gluckman et al. 2001). This analogy is strengthened by experimental observations that suggest that the cortex behaves like an excitable system (Sect. 3) and that the subcortex contains nuclei with diffuse projections that can inhibit or promote seizure spread. The main difference between the double sphere model for seizure spread and the control of spatiotemporal chaos in physical and chemical systems is that in the brain there are several different pathways for conduction each having different conduction velocities. The recent development of diffusion tensor imaging gives the possibility of measuring the white matter tracts that interconnect different regions of cortex and the cortex with the sub-cortex in patients with epilepsy (Beaulieu and Allen 1994; Rugg–Gunn et al. 2001). Although this method presently uses measurements of water molecule diffusivity to determine the orientation of these fiber tracts, it is conceivable that measurements of water mobility could provide even more information (see,

for example (Milton and Galley 1986)). In this context, the object would be to explain the observed patterns of intra-electrode coherence measured using subdural electrode grids on the basis of the properties of cortex, sub-cortex and their connection pathways in order to design startegies to prevent, limit, or abort the spread of seizures (Milton and Jung 2002).

References

Ajmone–Marsan C (1972). Focal electrical stimulation. In: Purpura DP, Penny JK, Tower D, Woodbury DM, Walker R (eds) *Experimental models of epilepsy: A manual for the laboratory worker*. Raven Press, New York, pp. 147–172.

Babb TL, Pretorius JK, Kupfer WR, Crandall PH (1989). Glutamate decarboxylase –immunoreactive neurons are preserved in human epileptic hippocampus. J. Neurosci. 9, 2562–2574.

Babb TL, Wilson CL, Isokawaakesson M (1987). Firing patterns of human limbic neurons during stereoencephalography (SEEG) and clinical temporal–lobe seizures. Electroencephal. Clin. Neurophysiol. 66: 467–482.

Baumgartner C, Lindinger G, Ebner A, Aull S, Serles W, Olbrich A, Lurger S, Czech T, Burgess R, Lüders H (1995). Propagation of interictal epileptic activity in temporal–lobe epilepsy. Neurology 45: 118–122.

Beaulieu C, Allen PS (1994). Water diffusion in the giant axon of the squid: Implications for diffusion–weighted MRI of the nervous system. Magn. Reson. Med. 31: 394–400.

Beaulieu C, Kisvarday Z, Somoygi P, Cynader M, Cowey A (1992). Quantitative distribution of GABA–immunopositive and –immunonegative neurons and synapses in the monkey striate cortex (Area 17). Cerebral Cortex 2: 295–309.

Beggs JM, Klukas J and Chen W (2007). Connectivity and dynamics in local cortical networks. This volume.

Beggs JM, Plenz D (2003). Neuronal avalanches in neocortical circuits. J. Neurosci. 23: 11167–11177.

Bragin A, Engel J Jr, Wilson CL, Fried I, Buzaki G (1999). High–frequency oscillations in human brain. Hippocampus 9: 137–142.

Bragin A, Mody I, Wilson CL, Engel J Jr (2002). Local generation of fast ripples in epileptic brain. J. Neurosci. 22: 2012–2021.

Breakspear M and Jirsa VK (2007). Neuronal dynamics and brain connectivity. This volume.

Bressloff PC, Coombes S (2002). Synchronization of synaptically coupled neural oscillators. In: Milton J, Jung P (eds) *Epilepsy as a dynamic disease*. Springer-Verlag, New York, pp. 83–114.

Buser P, Bancaud J (1983). Unilateral connections between amygdala and hippocampus in man: A study of epileptic patient with depth electrodes. Electroencephal. Clin. Neurophysiol. 55: 1–12.

Buser P, Bancaud J, Talairach J (1972). Electrophysiological studies of the limbic system with multiple multilead stereotaxic electrodes in epileptic patients. In: Somjen GC (ed). *Neurophysiology studied in man*. Excerpta Medica, Amsterdam, pp. 112–125.

Buzsáki G (1986). Hippocampal sharp waves: Their origin and significance. Brain Res. 398: 242–252.

Buzsáki G (1989). A two–stage model of memory trace formation: A role for "noisy" brain sites. Neuroscience 31: 551–570.

Buzsáki G, Czopf J, Kondakor I, Bjorklund A, Gage FH (1987). Cellular activity of intracerebrally transplanted hippocampus during behavior. Neuroscience 22: 871–883.

Buzsáki G and Traub RD (1997). Physiological basis of EEG activity. In: Engel J, Pedley TA (eds). *Epilepsy: A comprehensive textbook, Volume 1*. Lippincott–Raven, New York, pp. 819–830.

Cabrera JL, Milton JG (2002). On–off intermittency in a human balancing task. Phys. Rev. Lett. 89: 158702.

Cabrera JL, Bormann R, Eurich C, Ohira T, Milton J (2004). State–dependent noise and human balance control. Fluctuation Noise Lett. 4: L107–L117.

Campbell SA (2007). Time delays in neural systems. This volume.

Cerf R, Hassan E, Ouasdad EHE, Kahane P (2004). Criticality and synchrony of fluctuations in rhythmical brain activity: pretransitional effects in epileptic patients. Biol. Cybern. 90: 239–255.

Chen JWY, O'Farrell AM, Toga AW (2000). Optical intrinsic signal imaging in a rodent seizure model. Neurology 55: 312–315.

Chervin RD, Pierce PA, Connors BW (1988). Periodicity and directionality in the propagation of epileptiform discharges across neocortex. J. Neurophysiol. 60: 1695–1713.

Chialvo DR, Bak P (1999). Learning from mistakes. Neuroscience 90: 1137–1148.

Chkhenkeli SA, Chkhenkeli IS (1997). Effect of Therapeutic Stimulation of Nucleus Caudatus on Epileptic Electrical Activity of Brain in Patients with Intractable Epilepsy. Stereotactic Funct. Neurosurg. 69: 221–224

Chkhenkeli SA (2002). Direct deep brain stimulation: First steps towards the feedback control of seizures. In: Milton J, Jung P (eds) *Epilepsy as a dynamic disease*. Springer-Verlag, New York, pp. 249–262.

Chkhenkeli SA, Milton J (2002). Dynamic epileptic systems versus static epileptic foci? In: Milton J, Jung P (eds) *Epilepsy as a dynamic disease*. Springer-Verlag, New York, pp. 25–36.

Chkhenkeli SA, Sramka M, Lortkipanidze GS, Rakviashili TN, Bregvadze ES, Magalashvili GE, Gagoshidze TS, Chkhenkeli IS (2004). Electrophysiological effects and clinical results of direct brain simulation for intractable epilepsy. Clin. Neurol. Neurosurg. 106: 318–329.

Chiel HJ, Beer RD (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. Trends Neurosci. 20: 553–557.

Chu PH, Milton JG, Cowan JD (1994). Connectivity and the dynamics of integrate–and–fire neural networks. Int. J. Bifurc. Chaos 4: 237–243.

Cole BJ (1991). Is anilmal behavior chaotic? Evidence from the activity of ants. Proc. Roy. Soc. London B 244: 253–259.

Cooper R, Winter AL, Crow HJ, Walter WG (1965). Comparison of subcortical, cortical and calp activity using chronically indwelling electrodes in man. Electroencephal. Clin. Neurophys. 18: 217–228.

Darvas F and Leahy RM (2007). Functional imaging of brain activity and connectivity with MEG. This volume.

Dominguez LG, Wennberg RA, Gaetz W, Cheyne D, Snead OC, Velazquez JLP (2005). Enhanced synchrony in epileptiform activity? Local versus distant phase synchronization in generalized seizures. J. Neurosci. 25: 8077–8084.

Ebersole JS, Milton J (2002). The electroencephalogram (EEG): A measure of neural synchrony. In: Milton J, Jung P (eds) *Epilepsy as a dynamic disease*. Springer-Verlag, New York, pp. 51–68.

Ebersole JS, Pacia SV (1996). Localization of temporal lobe foci by ictal EEG patterns. Epilepsia 37: 386–399.

Emerson RG, Turner CA, Pedley TA, Walczak TS, Fergione M (1995). Propagation patterns of temporal spikes. Electroencephal. Clin. Neurophysiol. 94: 338–348.

Engel J Jr (1990). Functional explorations of the human epileptic brain and their therapeutic implications. Electroencephal. Clin. Neurophysiol. 76: 296–316.

Eurich CW, Hermann JM, Ernst UA (2002). Finite–size effects of avalanche dynamics. Phys. Rev. E 66: 066137.

Farley BG (1965). A neuronal network model and the "slow potential" of electro–physiology. Comp. Biomed. Res. 2: 265–294.

Ferree TC, Nunez PL (2006). Primer on electroencephalography for functional connectivity. This volume.

Fountas KN, Smith JR, Murro AM, Politsky J, Park YD, Jenkins PD (2005). Implantation of a closed–loop stimulation in the management of medically refractory fiocal epilepsy: A technical note. Stereo Funct. Neurosurg. 83: 153–158.

Fox SS, Norman RJ (1968). Functional congruence: An index of neural homogeneity and a new measure of brain activity. Science 159: 1257–1259.

Friedrich R, Fuchs A, Haken H (1991). Spatio–temporal EEG patterns. In: Haken H, Köpchen HP (eds) *Synergetics of rhythms*. Springer–Verlag, Berlin, pp. 119–128.

Fuchs A (2007). Beamforming and its applications to brain connectivity. This volume.

Fuchs A, Friedrich R, Haken H, Lehmann D (1987). Spatio–temporal analysis of multichannel EEG patterns. In: Haken H (ed) *Computational systems – natural and artificial*. Springer–Verlag, Berlin, pp. 74–83.

Gluckman BJ, Nguyen H, Weinstein SL, Schiff SJ (2001). Adaptive electric field control of epileptic seizures. J. Neurosci. 21: 590–600.

Gotman J (1983). Measurement of small time differences between EEG channels: Method and application to epileptic seizure propgation. Electroencephal. Clin. Neurophysiol. 56: 501–514.

Guyton AG (1976). *Textbook of Medical Physiology*. W. B. Saunders, Philadelphia.

Haglund MM (1997). Optical imaging. In: Engel J, Pedley TA (eds). *Epilepsy: A comprehensive textbook, Volume 1*. Lippincott–Raven, New York, pp. 1073–1079.

Haglund MM, Ojemann GA, Hochman DW (1992). Optical imaging of epileptiform activity from human cortex. Nature (London) 358: 668–671.

Haas KZ, Sperber EF, Moshe SL, Stanton PK (1996). Kainic acid–induced seizures enhance dentate gyrus inhibition by downregulation of GABA(B) receptors. J. Neurosci. 16: 4250–4260.

Horwitz B and Husain FT (2007). Simulation framework for large–scale brain systems. This volume.

Huang X, Troy WC, Yang Q, Ma H, Laing CR, Schiff SJ, Wu J–Y (2004). Spiral waves in disinhibited mammalian neocortex. J. Neurosci. 24: 9897–9902.

Iijima T, Witter MP, Ichikawa M, Tominaga T, Kajiwara R, Matsumoto G (1996). Entorhinal–hippocampal interactions revealed by real–time imagine. Science 272: 1176–1179.

Jentzen KJ and Kelso JAS (2007). Neural coordination dynamics of human sensorimotor behavior: A review. This volume.

Jirsch JD, Urrestarazu E, LeVan P, Olivier A, Dubeau F, Gotman J (2006). High–frequency oscillations during focal seizures. Brain 129: 1593–1608.

Kamiński M (2007). Multichannel data analysis in biomedical research. This volume.

Kelso JAS (1995). *The self–organization of brain and behavior*. The MIT Press, Cambridge, Ma.

Konigsmark BW, Abdullah AF, French JD (1958). Cortical spread of after–discharge in the monkey. Electroenceph. Clin. Neurophysiol. 10: 687–696.

Kreindler A (1965). *Experimenal epilepsy*. Elsevier, New York.

Ktonas PY, Mallart R (1991). Estimation of time–delay between EEG signals for epileptic focus localization – statistical error considerations. Electroencephal. Clin. Neurophysiol. 78: 105–110.

Kuebler D, Tanouye MA (2000). Modifications of seizure susceptibility in Drosphila. J. Neurophysiol. 83: 998–1009.

Langdon CG (1990). Computation at the edge of chaos: Phase transitions and emergent computation. Physica D 42: 12–37.

Lasota A, Mackey MC (1994). *Chaos, fractals, and noise: Stochastic aspects of dynamics*. Spriner-Verlag, New York.

Lesser RP, Kim SH, Beyderman L, Miglioretti DL, Webber WRS, Bare M, Cysyk B, Krauss G, Gordon B (1999). Brief bursts of pulse stimulation terminate afterdischarges caused by cortical stimulation. Neurology 53: 2073–2081.

Linkenkaer–Hansen K, Nikouline VV, Palva JM, Llmoniemi RJ (2001). Long–range temporal correlations and scaling behavior in human brain oscillations. J. Neurosci. 21: 1370–1377.

Lüders H (2004). *Deep brain stimulation and epilepsy*. Martin Dunitz, New York.

Matsumoto M, Saito S, Ohmine I (2002). Molecular dynamics simulation of the ice nucleation and growth process leading to water freezing. Nature (London) 416: 409–413.

Mihaliuk E, Sakurai T, Chirila R, Showalter K (2002). Feedback stabilization of unstable propagating waves. Phys. Rev. E 65: 065602.

Milton J (1996). *Dynamics of small neural populations*. American Mathematical Society, Providence, Rhode Island.

Milton JG (2000). Epilepsy: multistability in a dynamic disease. In: Walleczek J (ed) *Self–organized biological dynamics & nonlinear control*. Cambridge University Press, New York, pp. 374–386.

Milton J (2002). Medically intractable epilepsy. In: Milton J, Jung P (eds) *Epilepsy as a dynamic disease*. Springer-Verlag, New York, pp. 1–14.

Milton J (2002). Insights into seizure propagation from axonal conduction times. In: Milton J, Jung P (eds) *Epilepsy as a dynamic disease*. Springer-Verlag, New York, pp. 15–23.

Milton JG, Chu PH, Cowan JD (1993). Spiral waves in integrate–and–fire neural networks. In: Hanson SJ, Cowan JD, Giles CL (eds) *Advances in neural information processing systemsm, Volume 5*. Morgan Kaufmann, San Mateo, pp. 1001–1007.

Milton JG, Galley WC (1986). Evidence for heterogeneity in DNA–associated solvent mobility from acridine phosphorescence spectra. Biopolymers 25: 1673–1684.

Milton JG, Gotman J, Remillard GM, Andermann F (1987). Timing of seizure recurrence in adult epileptic patients. Epilepsia 28: 471–478.

Milton J, Jung P (2002). *Epilepsy as a dynamic disease*. Springer–Verlag, New York.

Milton J, Mundel T, an der Heiden U, Spire J-P, Cowan J (1995). Traveling activity waves. In: Arbib MA (ed) *The handbook of brain theory and neural networks*. MIT Press, Cambridge, MA, pp. 994–997.

Milton JG, Foss J, Hunter JD, Cabrera JL (2004). Controlling neurological diseases at the edge of stability. In: Pardaols PM, Sackellaris JC, Carney PR, Iasemidis LD (eds). *Quantitative neurosciences: Models, algorithms, diagnostics, and therapeutic devices*. Kluwer Academic Publishers, Boston, pp. 117–143.

Milton JG, Mackey MC (2000). Neural ensemble coding and statistical periodicity: Speculations on the operation of the mind's eye. J. Physiol. (Paris) 94: 489–503.

Moreau L, Sontag E (2003). Balancing at the border of instability. Phys. Rev. E 68: 020901

Motamedi GK, Lesser RP, Miglioretti, Mizuno–Matsumoto Y, Gordon B, Webber WRS, Jackson DC, Sepkuty JP, Crone NE (2002). Optimizing parameters for terminating cortical afterdischarges with pulse stimuation. Epilepsia 43: 836–846.

Netoff TI, Schiff SJ (2002). Decreased neuronal synchronization during experimental seizures. J. Neurosci. 22: 7297–7307.

Nunez PL (1981). *Electric fields of the brain: The neurophysiology of EEG*. Oxford University Press, New York.

Nunez PL (1995). *Neocortical dynamics and human EEG rhythms*. Oxford University Press, New York.

Nunez PL, Srinivasan R (2006). A theoretical basis for standing and traveling brain waves measured with human EEG with implications for an integrated consciousness. Clin. Neurophysiol. 117: 2424–2435.

Pacia SV, Ebersole JS (1997). Intracranial EEG substartes of scalp ictal patterns from temporal lobe foci. Epilepsia 38: 642–654.

Parish LM, Worrell GA, Cranstoun SD, Stead SM, Pennell P, Litt B (2004). Long–range temporal correlations in epileptiform and non–epileotgenic human hippocampus. Neuroscience 125: 1069–1076.

Penfield W, Jasper HH (1954). *Epilepsy and the Functional Anatomy of the brain*. Little Brown, Bostoin.

Petsche H, Pockberger H, Rappelsberger P (1988) Cortical structure and electrogenesis. In: Basar E (ed) *Dynamics of sensory and cognitive processing by the brain*. Springer–Verlag, New York, pp 123–129.

Petsche H, Prohaska O, Rappelsberger P, Vollmer R, Kaiser A (1974). Cortical seizure patterns in multidimensional view: Information–content of equipotential maps. Epilepsia 15: 439–463.

Pikovksy A, Rosenbaum M and Kürths J (2003). *Synchronization: A universal concept in nonlinear sciences*. Cambridge University Press, Cambridge, Ma.

Rinzel J, Terman D, Wang X-J, Ermentrout B (1998). Propagating patterns in large-scale inhibitory neuronal networks. Science 279: 1351–1355.

Romer AS (1967). *The Vertebrate Body, 3rd edition*. W. B. Saunders, Philadelphia.

Rosenblum M, Pikovsky A, Kurths J (1996). Phase synchronization of chaotic oscillators. Phys. Rev. Lett. 76: 1804–1807.

Lopes da Silva F, van Rotterdam A (1982). Biophysical aspects of EEG and magnetoencephalogram generation. In: Niedermeyer E, Lopes da Silva F (eds). *Electro–encephalography: Basic principles, clinical applications, and related fields*. Urban & Schwarzenberg, Baltimore, pp. 29–41.

Rugg–Gunn FJ, Eriksson SH, Symms MR, Barker GJ, Duncan JS (2001). Diffusion tensor mapping of cryptogenic and acquired partial epilepsies. Brain 124: 627–636.

Sakurai T, Mihaliuk E, Chirila F, Showalter K (2002). Design and control of wave propagation patterns in excitable media. Science 296: 2009–2012.

Schiff SJ, Colella D, Jacyna GM, Hughes E, Creekmore JW, Marshall A, Bozek–Kuzmicki M, Benke G, Gaillard WD, Conry J, Weinstein SR (2000). Brain chirps: spectrographic signatures of epileptic seizures. Clin. Neurophysiol. 111: 953–958.

Segev R, Benveniste M, Hulata E, Cohen N, Palevski A, Kapon E, Shapira Y, Ben–Jacob E (2002). Long term behavior of lithographically prepared in vitro neuronal networks. Phys. Rev. Lett. 88: 118102.

So P, Barreto E, Josić K, Sander E, Schiff SJ (2002). Limits to the experimental detection of nonlinear synchrony. Phys. Rev. E 65: 046225.

Steriade M, Contreras D (1998). Spike–wave complexes and fast components of cortically generated seizures. I. Role of neocortex and thalamus. J. Neurophysiol. 80: 1439–1455.

Steriade M, Amzica F (1994). Dynamic coupling among neocortical neurons during evoked and spontaneous spike–wave seizure activity. J. Neurophysiol. 72: 2051–2069.

Stevens CF (1965). How cortical interconnectedness varies with network size. Neural Comp. 1: 473–479.

Sutherling WW, Barth DS (1989). Neocortical propagation in temporal–lobe spike foci on magnetoencephalography and electroencephalography. Ann. Neurol. 25: 373–381.

Sutula T, Cascino G, Cavazos J, Paranda I (1989). Mossy fiber synaptic reorganization in the epileptic human temporal–lobe. Ann. Neurol. 26: 321–330.

Tao JX, Ray A, Hawes–Ebersole S, Ebersole JS (2005). Intracranial EEG substartes of scalp EEG interictal spikes. Epilepsia 46: 669–676.

Towle VL, Ahmad F, Kohrman M, Hecox K, Chkhenkeli S (2002). Electrocorticographic coherence patterns of epileptic seizures. In: Milton J, Jung P (eds) *Epilepsy as a dynamic disease*. Springer-Verlag, New York, pp. 69–81.

Traub RD, Miles R (1991). *Neuronal networks of the hippocampus*. Cambridge University Press, Cambridge, Ma.

Traub RD, Miles R, Wong RKS (1989). Model of rhythmic population oscillation in the hippocampal slice. Science 243: 1319–1325.

Traub RD, Whittington MA, Buhl EH, LeBeau FE, Biddig A, Boyd S, Cross H, Baldeweg T (2001). A possible role for gap junctions in generation of very fast EEG oscillations preceding the onset of, and perhaps initiating, seizures. Epilepsia 42: 153–170.

van Drongelen W, Lee HG, Hereld M, Chen ZY, Elsen FP, Stevens RL (2005). Emergent epileptiform activity in neural networks with weak excitatory synapses. IEEE Trans. Neural Sys. Rehab. Engn. 13: 236–241.

Weir B (1964). Spikes–wave from stimulation of reticular cortex. Arch. Neurol. 11: 209–218.

Winfree AT (2002). Are cardiac waves relevant to epileptic wave propagation? In: Milton J, Jung P (eds) *Epilepsy as a dynamic disease*. Springer-Verlag, New York, pp. 165–188.

Worrell GA, Cranstoun SD, Echauz J, Litt B (2002). Evidence for self–organized criticality in human epileptic hippocampus. Neuroreport 13: 2017–2021.

Worrell GA, Parish L, Cranstoun SD, Jonas R, Baltuch G, Litt B (2004). High–frequency oscillations and seizure generation in neocortical epilepsy. Brain 127: 1496–1506.

Wright JJ (1999). Simulation of EEG: dynamic changes in synaptic efficacy, cerebral rhythms, and dissipative and generative activity in cortex. Biol. Cybern. 81: 131–147.

Wyler AR, Ojemann GA,Ward AA Jr (1982). Neurons in human epileptic cortex: Correlation between unit and EEG activity. Ann. Neurol. 11: 301–308.

Large Scale Brain Dynamics and Connectivity in the Minimally Conscious State

Nicholas D Schiff

Weill Medical College of Cornell University, 1300 York Avenue,
New York, 10021

It is increasingly recognized that some patients who sustain severe brain injuries nonetheless recover the ability to communicate and express cognitive function after long convalescent intervals. The physiological mechanisms underlying this recovery are poorly understood and few measurements are aimed at identifying the potential for further recovery in individual patients. In this chapter we review the emerging literature of investigative studies that employ quantitative measurements of cerebral function in the minimally conscious state (MCS), the clinical subcategory from which patients have made late recoveries of goal-directed behavior and communication. Several studies demonstrate evidence of intact large-scale cortical networks in MCS patients providing a potential substrate for both behavioral fluctuations and residual cognitive capacities. Thus, MCS presents several challenges for understanding the role of brain connectivity and dynamics underlying integrative functions of the human brain.

Based on findings from our own studies and others in the published literature we develop a working hypothesis that MCS reflects the preservation of large-scale cerebral cortical networks in a chronically underactive brain. Preliminary studies using quantitative electroencephalography (EEG), fluorodeoxyglucose-positron emission tomography (FDG-PET), diffusion tensor imaging (DTI) and functional magnetic resonance imaging (fMRI) techniques provide support for this model. In some patients MCS may reflect a brain retaining recruitable, functionally connected networks that exhibit unstable thalamocortical dynamics producing bi-hemispheric dysfunction. Several possible physiologic mechanisms may limit more normal expression of network function in MCS patients including hypersynchrony within selective forebrain networks, broad loss of tonic background activity across cerebral structures, and specific functional disturbances arising from focal injuries to midline brain structures. We propose a strategy of combined measurements to isolate such possible contributions to global brain dysfunction in MCS.

Spontaneous emergence late in the course of MCS may also depend in some cases not only on changes in brain dynamics but also on structural changes

that arise over long time courses. Mechanisms and data suggestive of such a slow variable relating to recovery are briefly reviewed.

We conclude that there is a need to further develop measurements of large-scale brain dynamics in MCS and longitudinally track changes in brain functional and structural connectivity associated with recovery. Points of contact with several other contributions to this volume are highlighted, as are directions for future research.

1 Framing the Problem

The minimally conscious state (MCS) denotes a specific level of functional recovery following severe brain injuries characterized by reproducible and unequivocal evidence of response to the environment (Giacino et al. 2002). MCS patients fail to recover the ability to communicate reliably, although they typically demonstrate spontaneous behavioral fluctuations, and rare patients intermittently exhibit sustained periods of consistent communication with later regression back into MCS (e.g., Burrus and Chacko 1999, Clauss et al. 2001). The possibility that this condition may produce a unique burden of isolation motivates efforts to better define brain function in MCS patients in the hope of promoting further recovery (cf. Fins 2003). Emergence from MCS is defined by the recovery of reliable communication. MCS patients near the borderline of emergence may harbor significant residual capacity for interactive awareness and communication.

MCS patients and other patients with severe disabilities following traumatic brain injury (TBI) constitute a large and significantly understudied patient population. In the United States there is an estimated prevalence of 3–5 million Americans with significant residual functional impairment following TBI (NIH Consensus Panel 1999). The prevalence of MCS *per se* is estimated at 112,000–280,000 adult and pediatric patients in the U.S. (Strauss et al. 2000). These considerations motivate the goal of identifying the mechanisms limiting cognitive capacity in MCS patients. Understanding mechanisms underlying MCS will also provide important insight into the recovery of cognitive function in patients with less severe brain injuries, as principles of brain network recovery are unlikely to be sharply divided across these descriptive functional categories.

Figure 1 places MCS in the context of other global disorders of human consciousness. It is important to distinguish MCS from coma and the vegetative state (VS) as well as the locked-in state (LIS), which is not a disorder of consciousness but may be mistaken for these other conditions. By definition coma is a condition in which patients show no response to environmental stimulation and cannot be aroused from an eyes closed state that resembles sleep. Comatose patients show distinct reflex movements of the eyes and limbs (Plum and Posner 1982). The vegetative state, VS, is also characterized by

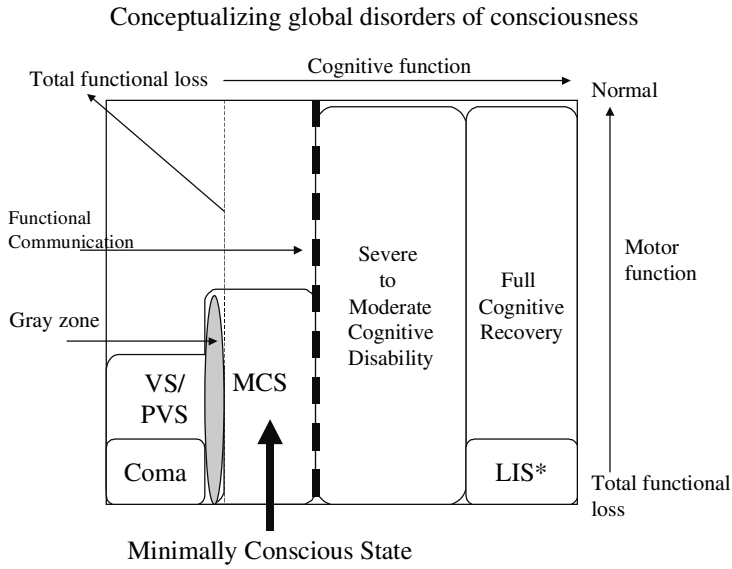


Fig. 1. Overview of global disorders of consciousness indexed by level of cognitive (*horizontal axis*) and motor functions (*vertical axis*). VS, vegetative state; MCS, minimally conscious state; LIS, locked-in state; asterisk indicates that LIS is not a disorder of consciousness. See text for further details. Figure adapted from MIT Press with permission)

a lack of response to environmental stimuli and is distinguished from coma by the recovery of a crude form of sleep-wake cycling with alternation of eyes open and eyes closed time periods. VS patients also may exhibit a variety of stereotyped (often limbic) behavioral displays (Jennett and Plum 1972). MCS patients, however, demonstrate clear and unequivocal responses to stimulation or spontaneous volitional behaviors (Giacino et al. 2002). MCS patients may sustain visual fixation, follow one-step commands, verbalize, and in some cases briefly communicate. Very limited behavioral responses may also be seen in locked-in patients (LIS) where incomplete motor responses may mask a normal level of conscious experience (cf. Laureys et al. 2005).

Spontaneous fluctuations in level of response are often seen in MCS patients and verified examples of late recoveries of consistent communication following severe brain injuries are drawn from this population. Brain mechanisms underlying such a cognitive reserve are unknown and measurements to predict recovery and functional level are currently unavailable. To frame these questions properly both empirical data and conceptual models are needed. Below we consider the available evidence of brain network activity and connectivity in MCS and possible mechanisms limiting the greater functional recovery in this patient population.

2 Brain Connectivity in MCS

Few pathological studies have described findings from MCS patients. Jennett and colleagues (2001) compared autopsies findings of TBI patient who remained in VS or with severe disability including 12 patients with histories consistent with MCS at the time of death. Patients remaining in VS showed extensive neuronal death throughout the thalamus of both cerebral hemispheres (cf. Adams et al. 2000). A majority of the patients in the severely disabled group showed focal brain injures, without diffuse axonal injuries or diffuse thalamic injuries. Of note, this pattern of widely preserved brain structure was seen in 2 of the MCS patients. The findings indicate a wide variance of underlying substrate of brain connectivity producing a low-level clinical exam; this provides an important starting point when considering the question of brain network activity after severe injuries.

Permanent VS can be modeled as arising in the context of a massively disconnected brain that preserves only few thalamocortical projections from the sensory periphery to the early cortical regions (see Schiff 2005 for review). Sufficient bilateral damage to primarily to the thalamus leads to permanent loss of the resting patterns of cortical activity associated with goal-directed behavioral and sensory responsiveness. Patients in a permanent VS show loss of late evoked potential components, and cerebral metabolic activity, as well as gross abnormalities of the EEG spectrum (cf. Schiff et al. 2002). This picture of VS gives way to several variations particularly if VS is not the result of widespread anoxic or traumatic brain damage.

Identification of diffuse structural injury has been explored as a method of outcome assessment in severe brain injuries producing VS using MRI indices. Some studies have found relatively high predictive accuracy ($\sim 84\%$) when combined with time in VS for a permanently vegetative outcome of TBI (cf. Kampfl et al. 1998). Recent observations, however, indicate that patients fulfilling these criteria can recover after long intervals if reaching MCS within one year. In our ongoing studies we have identified two MCS patients with very severe diffuse axonal injury patterns one who emerged at 8 months and the other 19 years after remaining in MCS (Voss et al. 2005 and unpublished observations). These observations underscore the complexity of the structure-function relationships that may arise in a severe damaged human brain over extended periods of time. Furthermore, experimental studies suggest the possibility for late and ongoing changes in brain structure to arise after severe injuries. Dancause et al. (2005) found extensive cortical rewiring in the adult primate brain after stroke lesions. In their studies done 5 months after injury, extensive proliferation of novel terminal fields within distant cortical states were produced by axonal sprouting. Several other mechanisms that could support cortical rewiring have also been identified (Chklovskii et al. 2004). These clinical and experimental findings indicate that ongoing assessment of both grey and white matter structures will be necessary to understand recovery from severe brain injuries. Longitudinal evaluations of changes

in brain structure using magnetic resonance spectroscopy (^{1}H-MRS) and diffusion tensor imaging (DTI) measurements will provide important information (cf. Lobaugh and Alexander, this handbook).

Beyond the contributions of multi-focal cortical or diffuse axonal patterns of injury an important consideration in the evaluation of structural injuries producing impaired consciousness is the integrity of the paramedian regions of the thalamus and midbrain (Schiff and Plum 2000). Sustained VS and MCS can arise from focal injuries concentrated bilaterally in these regions (see e.g., patient 5 in Schiff et al. 2002). Such injuries will typically encompass damage to the mesencephalic reticular formation and projections into the intralaminar thalamic nuclei and often occur following brain swelling from injury because of the vulnerability of these structures to be pressed through the base of the skull (cf. Plum and Posner 1982). Most severe brain injuries present a mix of these pathologies, combining moderately diffuse or multifocal regional damage with limited injuries to paramedian structures. The functional role of the neuronal populations within the paramedian thalamus and mesencephalon in MCS is considered below (see Schiff and Plum 2000, Schiff and Purpura 2000 for review).

3 Large Scale Brain Network Function in MCS

Understanding the time-evolving functional recovery of large scale brain dynamics in the setting of severe brain injury will provide important clues to the mechanisms underlying human consciousness and cognitive capacities. At present only a small number of MCS patients have been studied using functional neuroimaging methods. Boly et al. (2004a) compared brain activation in response to simple auditory stimuli from 5 MCS patients using functional ^{15}O PET techniques earlier applied by Laureys et al. (2000) to study 15 VS patients. In this study, both the MCS patients and healthy controls both activated auditory association regions in the superior temporal gyrus that did not activate in the PVS patients. In contrast all groups activated primary auditory sensory regions consistent with evoked potential studies and other functional imaging of VS patient responses (see Laureys et al. 2004, Schiff 2005 for reviews of neuroimaging in VS patients). In addition, Boly et al. found a strong correlation of the auditory cortical region response with that of frontal cortical regions possibly suggesting cerebral processing associated with higher-order integrative function. Similar findings of a widely preserved normal pain network response were also observed following noxious somatosensory stimulation in the MCS patients (Boly et al. 2005).

Coleman et al. (2005) used combined FDG-PET and quantitative EEG studies to investigate homeostatic relationships between regional cerebral metabolic rates and the distribution of EEG power across frequencies. In a small sample of VS and MCS patient, coupling between neuronal electrical function and cerebral metabolism was preserved in the MCS but not the VS

patients. This finding adds important detail to the well-established correlation of very low resting metabolism ($< 50\%$ of normal) in VS (e.g., Levy et al. 1987).

Schiff et al. (2005) studied two MCS patients using passive fMRI paradigms and FDG-PET. FDG-PET measurements demonstrated severe reductions in resting brain metabolic activity to $\sim 50\%$ normal, roughly equivalent to the low levels found in FDG-PET studies of VS patients (discussed above). Brain activation patterns from the patients and 7 control subjects were obtained using passive language stimulation fMRI paradigms similar to those in used in normal subjects and neurosurgical candidates to map language networks (Hirsch et al. 2000). Two 40-second narratives were pre-recorded by a familiar relative and presented as both normal speech and played as time-reversed. Forward presentations generated robust activity in several language-related areas in both patients. Figure 2 shows data corresponding to one MCS patient. Significant regions of activation included the inferior and medial frontal gyri (not shown), superior and middle temporal gyri, as well as primary and secondary visual areas including the calcarine sulcus, inferior and middle occipital gyri, precuneus, cuneus, and inferior parietal lobule. For this patient, the total volume and the specific regional cerebral activity overlapped with controls for the forward narrative presentations. Importantly, broad network activations in the two MCS patients were only seen for forward speech, and not for reversed speech; in normal controls, both presentations activated similar areas. Additionally, the posterior occipital activity suggests possible visualization during presentation of the forward narratives. While this study clearly demonstrates language-specific activations of large-scale networks in MCS patients, the implications of this difference between MCS patients and normals are unclear. One possibility is that the lack of activation for the time-reversed narratives reflects the loss of anticipatory, ongoing perceptual processing of environmental stimuli in MCS patients. The time-reversed narratives may thus fail to overcome low resting activity in the MCS patients (inferred from the low metabolic rates measured using PET) and engage dormant large-scale network processing capacities.

Other investigators have also identified language responsive networks in MCS patients. Bekinschtein et al. (2004) reported brain activations obtained using fMRI in a MCS patient recovering from traumatic brain injury. Comparison of responses to presentations of the patient's mother's voice and a neutral control voice revealed selective activation of the amygdala and insular cortex suggesting emotional processing associated with the mother's voice. Laureys et al. (2004) identified more widespread activation of cerebral networks with emotionally salient stimuli in an MCS patient. The demonstration of recruitable large-scale networks in MCS patients importantly establishes a continuity of functional correlations of cognition in this patient population and normal subjects. The findings thus focus attention on improving our measurements of the quality of network interactions in MCS brain and other conditions associated with less severe brain damage.

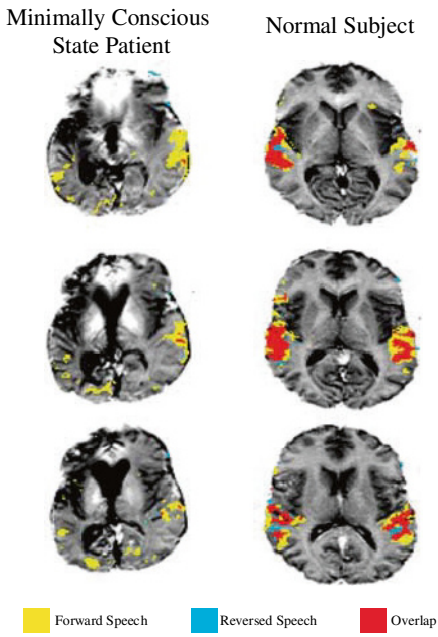


Fig. 2. Brain activations in response to passive language stimuli from a MCS patient and normal measured by functional magnetic resonance imaging. Yellow color indicates response to spoken narratives, blue color indicates response to time-reversed narratives, red color indicates regions of overlapping response to both conditions. See text for details of paradigm. Figure adapted from Schiff et al. 2005 *Neurology* (with permission)

As Owen et al. (2005a) have recently proposed, more hierarchically structured evaluation of network function will be necessary to interrogate network processing in MCS and VS patients. In a recent single case study, Owen and colleagues found evidence for processing of semantic information in a patient with an exam consistent with VS. Related work by Kotchoubey (2005) and colleagues has examined EEG and evoked response in large numbers of VS and MCS patients and found differential response profiles correlated with the spectral content of the background EEG rhythms. In these studies MCS and some VS patient also showed differential responses to semantic content of auditory stimuli.

In the aggregate, available studies suggest that MCS has a broad spectrum pathologically that includes some patients with strongly connected brain networks capable, in principle, of higher-level cognitive processing, including sufficient resources to regain access to communication, memories and internally motivated goal-directed behaviors. These findings motivate further research to understand the longitudinal process of recovery of brain network function in MCS and the mechanisms limiting the expression of this capacity, if present, in individual patients.

4 A Working Hypothesis for the Role of Altered Brain Dynamics Underlying the Minimally Conscious State

The studies reviewed above indicate that MCS is not generally characterized by global structural disconnection of the brain nor functionally limited to early sensory processing as measured by functional imaging studies or evoked potentials. However, despite the evidence of distributed cerebral activity in fMRI or PET studies and electrophysiological evidence of processing of auditory stimuli including aspects of semantic encoding, MCS *per se* reflects a failure of consistent goal directed behavior and communication. Thus, these data challenge us to reconcile the clear behavioral limitations of the patients with the evidence of, in some cases, normal network activation profiles. Setting aside the patients for whom a specific loss of motor function is predicted by structural injuries, at least two general hypotheses can be supported by the current literature (see Schiff 2005 more detailed review). On the one hand, the lack of self-generated behaviors may reflect the loss of intrinsic internal states characterized by specific, if nonetheless currently underspecified, brain dynamics. A complementary mechanism likely underlying failure of consistent engagement of potentially functional brain networks in MCS is the presence of stable abnormal brain dynamics within the thalamocortical system impairing intra- and interhemispheric interactions. To address the possible contributions of these mechanisms it is necessary to examine the state-dependence of brain network activations. That this is necessary is further supported by evidence of significant variations in observed clinical exam of MCS patients depending on the time in the day or the patient's recumbent posture among other variables (Elliott et al. 2005).

In both MCS patients studied by Schiff et al. (2005) with preserved large-scale network activations demonstrated by fMRI and low global resting metabolic rates assessed by FDG-PET, quantitative EEG analyses revealed hemispheric abnormalities of inter-regional coherence that corresponded to differences in resting metabolism (Kobylarz et al. 2005). Coherence is a measure of cross-correlation in the frequency domain (Mitra and Pesaren 1999). A statistically significant coherence suggests potential relationships between two given signals, e.g., one driving the other, mutual driving, or both partly driven by a common input signal (cf. Bendat and Piersol 2000). Because coherence is a ratio of coherent power to total power, changes in coherence cannot be simply the result of amplification or filtering of the power spectrum, but rather imply changes in functional connectivity or increased common input. The resting awake EEG studies in both MCS patients revealed significant reductions in inter-regional coherence of the less metabolically active hemisphere (see Fig. 3 for an example). The abnormalities of EEG coherence were apparent in all adjacent electrode pairs of the less metabolically active hemisphere indicating a significant alteration of the functional integration of cortical regions . For the patient the shown in Fig. 3, this interregional coherence pattern had a marked dependence on arousal state with sharp coherence decreases ob-

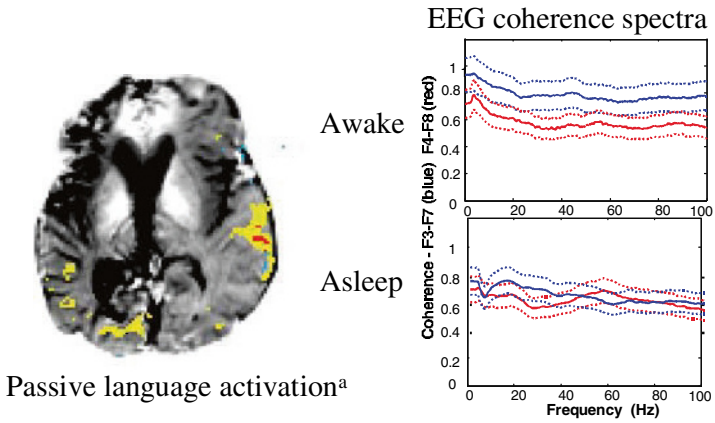


Fig. 3. Multimodal imaging data from MCS patient. Upper Left: single image of passive language fMRI activations, see Fig. 2 [image from *Neurology* with permission]; Upper Right: EEG Coherence spectra from adjacent frontal EEG leads (F3-F7 [*left hemisphere, blue*], F4-F8 [*right hemisphere, red*]). Coherence spectra demonstrate a sharp, broadband reduction of coherence over the right frontal lobe channels during awake states with near total normalization of the difference between the inter-regional coherence in the two hemispheres during sleep. Dotted lines indicate 95% confidence levels [EEG coherence spectra and resting metabolic data from Kobylarz et al. 2005, unpublished images]

served across frequencies only in the state of wakefulness (see normalization in the sleep state). These quantitative EEG findings suggest that the abnormalities observed in the coherence measurements obtained during wakefulness could be a possible marker of abnormal dynamics masking greater network response (Kobylarz et al. 2005).

Although not assessed in these patients, it is possible that salient stimuli such as personally meaningful narratives may produce a relative normalization of both resting metabolism and interregional EEG coherence. Several studies have addressed EEG power spectral and coherence measures in TBI patients and other forms of brain injury or dysfunction (Knyazeva and Innocenti 2001, Walker et al. 2000, Thatcher et al. 1998, 2001). Thatcher et al. (2001) reported an EEG severity index in which EEG coherence, phase and amplitude differences but not power variables provided significant contributions. These findings support the sensitivity of spectral measures for identifying patho-

physiological relevant changes in the EEG of MCS patients and suggest the potential for further development. The intermittent fluctuations of cognitive function observed in MCS patients over periods of hours, days or weeks, are proposed to originate from unstable interactions of the brainstem arousal systems with paramedian regions of the midbrain and thalamus that support activation and gating of large-scale cerebral networks (Schiff et al. 2005, Schiff 2005). The broadband, hemispheric reductions in EEG coherence observed in MCS patients may directly reflect functional impairment of thalamic driving inputs to the cerebral cortex.

We speculate that fluctuations of interactive behavior observed clinically in the MCS patient shown in Fig. 3 relate to ongoing dynamic changes of interregional interactions within and across the hemisphere controlled by such a thalamic driving input. The FDG-PET data shown Fig. 3 provide support for this hypothesis. As seen in the figure, the right thalamus shows markedly reduced metabolic activity throughout its visible extent. Schiff et al. (2005) proposed that impaired thalamocortical function resulting from bilateral damage to the thalami and upper brainstem during the acute phase of injury provided the mechanism for low metabolic activity, and an inferred loss of ongoing activation of the brain networks in both patients. Selective structural injuries to the paramedian thalamus and midbrain are unique in producing hemisphere-wide metabolic reductions (Szelies et al. 1991, Caselli et al. 1991). In addition, alteration of global dynamics may arise through structural injuries to these same regions (reviewed in Schiff 2005)

Taken together, the correlation of metabolic depression in the thalamus and broadband reductions of coherence across the entire cerebral hemisphere suggest a basis for a bi-hemispheric dysfunction in the MCS brain: an overall loss of cerebral integrative activity in one hemisphere that remains nonetheless functionally connected to the less inactive hemisphere may prevent organized responses of large-scale networks across the hemispheres. For example, the normalization of coherence patterns in during sleep may model similar fluctuations of hemispheric integration within the state of wakefulness when hemispheric is synchronized by salient stimuli as opposed to generic changes in the ongoing sleep-wake cycle.

Based on the above, we hypothesize that in some patients MCS reflects a chronically underactive brain that nonetheless retains widely connected and recruitable cerebral networks capable of supporting sensorimotor integration, limited behaviors, and intermittent communication. Testing this hypothesis will require establishing quantitative measurements to evaluate the correlation of quantitative EEG measurements with metabolic, structural and functional activation measures. It is expected that MCS patients will exhibit severely depressed global cerebral metabolic rates measured by FDG-PET reflecting marked downregulation of neuronal activity. The spectral content of the EEG may covary with metabolic rates in the thalamus and cerebral hemisphere. Changes of these measures in response to structured stimuli could index changes in the integrative activity of cerebral networks within and across brain

hemispheres in MCS (as suggested by the data shown in Fig. 3). If behavioral variations correlate with specific patterns of EEG abnormalities, a recruitable normalization of EEG responses may then be a marker for cognitive reserve in MCS.

An important limitation of the above hypothesis, however, is that the wide variance of possible brain injuries producing a clinical exam consistent with MCS limits any one model from having application in all cases. Other MCS patients who maintain only modest evidence of sensorimotor integration may not harbor extensive responsive brain networks. Moreover, even recovery of goal-directed communication and behavior following MCS is likely to be accompanied by significant cognitive disability. However, since recovery of the capacity to communicate and organize behavior presents an important threshold for many families and different ethical obligations to the patients themselves (cf. Fins 2005), it seems warranted to apply further efforts to develop measurements to identify patients who harbor greater network reserve.

5 Late Emergence from MCS

Perhaps no aspect of the problem of impaired human consciousness captures more general interest than the rare occasions of late recovery of communication and goal-directed behavior following severe brain injury. Although several verifiable cases have arisen, almost none have been evaluated with brain imaging and reported in the scientific literature. In a collaborative research effort we have recently extensively characterized brain structural connectivity and functional activation profiles in a 40 year-old male who fully recovered expressive and receptive language after remaining in MCS for 19 years following a severe traumatic brain injury (Voss et al. 2005, Goldfine et al. 2005; Giacino, Hirsch, Schiff unpublished data). Structural brain imaging using standard MRI evaluations showed marked volume loss and brainstem atrophy; diffusion tensor imaging (DTI) indicated very severe diffuse axonal injury with for example, a reduction of volume of the corpus callosum to one-third or two-thirds normal volume throughout the structure and aggregate measures of diffusion showed marked abnormality compared to normal subjects (Voss et al. 2005). The patient's recovery was spontaneous and has been sustained for over 2 years with further gradual clinical improvements.

Evaluation of brain network activations using the passive language paradigm (cf. Fig. 2) in this patient showed a response consistent with that of normal subjects for both forward and time-reversed narratives. EEG responses to the presentation of the same auditory stimuli were also obtained from multiple presentations during resting wakeful states over several days (Goldfine et al. 2005). Spectral analysis of averaged EEG from single channels across three conditions (forward narratives, time-reversed narratives and baseline) revealed changes in EEG leads over the left and right temporal lobes. These EEG

channels showed suppression of activity in the ∼0–10Hz frequency range and increases in power in the ∼30–50 Hz frequency range during presentation of passive 'forward' language stimuli compared to silent baseline periods. The time-reversed narratives produced a similar but not identical profile across the temporal lobe leads. The findings are consistent with other studies of electrocorticographic recordings from the human temporal lobe during language processing (Crone et al. 2001). The temporal lobe response profiles correlated with measured increases in BOLD response from the patient's temporal cortices underlying these electrode positions as identified in independent fMRI experiments using the same stimuli (J. Hirsch unpublished). Initial pilot EEG studies in a normal human subject's showed similar response profiles. Taken together, the fMRI and EEG data reflect a strong preservation of normal brain dynamics further raising the question of how this patient's recovery occurred.

One possible clue is that in addition to severe reduction of brain connectivity as assessed with DTI, the patient showed unusual large regions of increased connectivity in posterior brain structures not seen in 20 normal subjects (Voss et al. 2005). These findings suggest the possibility of a slow variable of structural change playing a role in the patient's recovery. Similar clinical observations in other MCS patients suggest such a slow process of recovery may exist and should be quantified through further structural imaging and longitudinal analysis of brain dynamics. McMillan and Herbert (2004) recently reported a 10-year follow-up on who remained in MCS for 2 years after a severe traumatic brain injury and then emerged to the point of indicating reliable yes or no responses. This patient, however, continued to recover 7–10 years following to a point of regaining the capacity to initiate conversation and express spontaneous humor and clear preferences. It is likely that such improvements over long time-intervals in severely brain-injured patients result from both functional and associated structural changes in the brain. Altering the level of neuronal activity may improve the probability that large-scale networks organize to support goal-directed behavior, perception, and communication. These functional changes may then lead to ongoing remodeling of brain structure over time.

6 Where do we go from here?

The small number of available datasets and the large variance of possible alterations of brain structure and dynamics arising in MCS prevent any more detailed conclusions. The studies reviewed above hopefully suggest the opportunities to link detailed brain measurements with simple models at this time and why it is potentially interesting. Many of the techniques and models discussed in this volume may ultimately play an important role. The significant structural injury patterns observed in MCS and other severely brain-injured patients provide a unique opportunity to examine predictions of physiologically realistic models such as presented by Breakspear and Jirsa (this volume). Preliminary attempts to relate measured human EEG to such

models appear promising (Rowe et al. 2004). Another important conceptual consideration is the role of specific neuronal structures that play critical roles in organizing network interactions. McIntosh (2004) has proposed that certain regions such as the hippocampus can be considered catalysts in determining network response profiles. Although the paramedian thalamus and midbrain are identified in this review as a common locus of brain damage in MCS, the catalyst approach is likely to identify the contributions of other cortical and subcortical structures to recovery of function following brain injuries.

What kind of new measurements will be useful? The measures applied above and suggested in this review remain relatively basic in terms of both brain imaging and signal processing techniques. However, this initial approach has the advantage of putting very complex brain injuries onto a more roughly equal footing than simple structure-function correlations. Eventually, more model driven approaches and paradigms will be needed. In addition to traditional methods of spectral analysis, related measures that allow more mechanistic inferences to be drawn from causal structure in the EEG data (see chapters by Kaminski, Stephan and Friston this volume) or related methods that model specific underlying generator structures (Repucci et al. 2001) may provide important and interpretable frameworks. A role for more general, abstract approaches to connectivity measurements as suggested by Tononi (this volume) and Sporns et al. (2005) is also supported by the EEG and DTI findings in the patient studied with late emergence from MCS. Tononi (2004) has proposed that consciousness *per se* requires a quantifiable level of complexity of connections; along these lines it would be interesting to examine recovery of EEG patterns against such metrics evaluating both slow structural alterations in the damaged brain and behavior. Ultimately, to usefully apply these methods to questions of recovery of brain function in MCS it will be necessary to obtain datasets from larger numbers of patients measured both prospectively and longitudinally.

Acknowledgments

We thank Dr. Joseph Fins for review. Work supported by NIH-NINDS, Charles A Dana Foundation, and James S. McDonnell Foundation grants.

References

Adams, J.H., Graham, D.I., Jennett, B. (2000) The neuropathology of the vegetative state after acute insult. Brain 123:1327–1338

Bekinschtein T, Leiguarda R, Armony J, Owen A, Carpintiero S, Niklison J, Olmos, L, Sigman L, Manes F.J (2004) Emotion processing in the minimally conscious state. J. Neurol Neurosurg Psychiatry 75(5): 788

Bendat, M, Piersol, J.S. (2000) Random Data, McGraw-Hill Publishers, New York

Boly M, Faymonville ME, Peigneux P, Lambermont B, Damas P, Del Fiore G, Degueldre C, Franck G, Luxen A, Lamy M, Moonen G, Maquet P, Laureys S. (2004) Auditory processing in severely brain injured patients: differences between the minimally conscious state and the persistent vegetative state. Arch Neurol. 61(2):233–8.

Boly, M, Faymonville ME, Peigneux P, Lambermont B, Damas P, Luxen A, Lamy M, Moonen G, Maquet P, Laureys S. (2005) Neuropsychological Rehabilitation 15:283–290

Burruss JW, Chacko RC. (1999) Episodically remitting akinetic mutism following subarachnoid hemorrhage. J Neuropsychiatry Clin Neurosci 11(1):100–2

Caselli, R.J., Graff-Radford, N.R., Rezai, K., (1991) Thalamocortical diaschisis: single-photon emission tomographic study of cortical blood flow changes after focal thalamic infarction. Neuropsych. Neuropsych. Behav. Neurol. 4, 193–214.

Chklovskii DB, Mel BW, Svoboda K. (2004) Cortical rewiring and information storage Nature. 431: 782–788

Clauss RP, van der Merwe CE, Nel HW. (2001) Arousal from a semi-comatose state on zolpidem. S Afr Med J. 91(10):788–9.

Coleman MR, Menon DK, Fryer TD, Pickard JD. (2005) Neurometabolic coupling in the vegetative and minimally conscious states: preliminary findings. J Neurol Neurosurg Psychiatry. 76(3):432–4.

Crone NE, Boatman D, Gordon B, Hao L. (2001) Induced electrocorticographic gamma activity during auditory perception. Clin Neurophysiol. 112(4):565–82.

Dancause, N, Barbay, S, Frost, SB, Plautz, EJ, Chen, D, Zoubina, EV, Stowe, AV and Nudo, RJ (2005) Extensive Cortical Rewiring after Brain Injury. J. Neurosci. 25 10167–10179.

Elliott L, Coleman M, Shiel A, Wilson BA, Badwan D, Menon D, Pickard J. (2005) Effect of posture on levels of arousal and awareness in vegetative and minimally conscious state patients: a preliminary investigation. J Neurol Neurosurg Psychiatry. 76(2):298–9.

Fins, JJ. (2003) Constructing an ethical stereotaxy for severe brain injury: balancing risks, benefits and access. Nat Rev Neurosci. 4(4):323–7.

Fins, JJ. (2005) Clinical pragmatism and the care of brain damaged patients: toward a palliative neuroethics for disorders of consciousness. Prog Brain Res. 150:565–82.

Giacino, J.T., Ashwal, S., Childs, N., Cranford, R., Jennett, B., Katz, D.I., Kelly, J.P., Rosenberg, J.H., Whyte, J., Zafonte, R.D., Zasler, N.D., (2002). The minimally conscious state: definition and diagnostic criteria. Neurology 58, 349–53.

Gold, L., Lauritzen, M., (2002) Neuronal deactivation explains decreased cerebellar blood flow in response to focal cerebral ischemia or suppressed neocortical function. Proc. Natl. Acad. Sci. 99, 7699–704.

Hirsch J, Ruge MI, Kim KH, Correa DD, Victor JD, Relkin NR, Labar DR, Krol G, Bilsky MH, Souweidane MM, DeAngelis LM, Gutin PH. (2000) An integrated functional magnetic resonance imaging procedure for preoperative mapping of cortical areas associated with tactile, motor, language, and visual functions. Neurosurgery. 47(3):711–21.

Jennett, B. (2002) The Vegetative State. Cambridge University Press.

Jennett B., Adams, J.H., Murray, L.S. et al. (2001) Neuropathology in vegetative and severely disabled patients after head injury. Neurology 56: 486–490.

Jennett, B and Plum, F. (1972) Persistent vegetative state after brain damage. A syndrome in search of a name.Lancet 1: 734–7

Kampfl A, Schmutzhard E, Franz G, Pfausler B, Haring HP, Ulmer H, Felber S, Golaszewski S, Aichner F. (1998) Prediction of recovery from post-traumatic vegetative state with cerebral magnetic-resonance imaging. Lancet 351(9118): 1763–7

Goldfine AM, Kobylarz, EJ, Hudson, AE, Giacino, JT, Hirsch, J and Schiff, ND (2005) Quantitative electroencephalography in a patient emerged from the minimally conscious state after 19 years. Society for Neuroscience 35^{th} Annual Meeting, Abstract

Knyazeva MG, Innocenti GM. J (2001) EEG coherence studies in the normal brain and after early-onset cortical pathologies. Brain Res Brain Res Rev 36(2–3): 119–28

Kobylarz, EJ, Hudson, AE, Kamal, A, DeBellis, R J, and Schiff, ND (2005) Power spectrum and coherence analysis of the electroencephalogram from two minimally conscious patients with severe asymmetric brain damage. Society for Neuroscience 35^{th} Annual Meeting, Abstract

Kotchoubey B, Lang S, Mezger G, Schmalohr D, Schneck M, Semmler A, Bostanov V, Birbaumer N. (2005) Information processing in severe disorders of consciousness: vegetative state and minimally conscious state. Clin Neurophysiol. 116(10):2441–53.

Laureys, S., Faymonville, M.E., Degueldre, C., Fiore, G.D., Damas, P., Lambermont, B., Janssens, N., Aerts, J., Franck, G., Luxen, A., Moonen, G., Lamy, M., Maquet, P., (2000a) Auditory processing in the vegetative state. Brain 123,1589–601.

Laureys S, Faymonville ME, Peigneux P, Damas P, Lambermont B, Del Fiore G, Degueldre C, Aerts J, Luxen A, Franck G, Lamy M, Moonen G, Maquet P. Cortical processing of noxious somatosensory stimuli in the persistent vegetative state. Neuroimage 2002; 17(2): 732–41.

Laureys, S.L., Owen, A.M., and Schiff, N.D. (2004) Brain function in coma, vegetative state and related disorders. Lancet Neurology 3(9):537–46.

Laureys S, Pellas F, Van Eeckhout P, Ghorbel S, Schnakers C, Perrin F, Berre J, Faymonville ME, Pantke KH, Damas F, Lamy M, Moonen G, Goldman S. (2005) The locked-in syndrome : what is it like to be conscious but paralyzed and voiceless? Prog Brain Res. 150:495–511.

Levy, DE, Sidtis, JJ, Rottenberg, DA, Jarden JO, Strother SC, Dhawan V, Ginos JZ, Tramo MJ, Evans AC, Plum F. (1987). Differences in cerebral blood flow and glucose utilization in vegetative versus locked-in patients. Ann Neurol *ogy* 22, 673–82

McIntosh ARJ (2004) Contexts and catalysts: a resolution of the localization and integration of function in the brain. Neuroinformatics. 2(2):175–82.

McMillan TM, Herbert CM. (2004) Further recovery in a potential treatment withdrawal case 10 years after brain injury. Brain Inj. 18(9):935–40.

Mitra P.P., Pesaren, B. (1999) Analysis of dynamic brain imaging data. Biophysical Journal 76(2):691–708

NIH Consensus Development Panel on Rehabilitation of Persons with Traumatic Brain Injury JAMA 1999, 282(10): 974–983

Owen, A, Coleman, M, Menon, D, Johnsrude, IS, Rodd, JM, Davis, MH, Taylor, K, Pickard, JD (2005) Residual auditory processing in persistent vegetative state: a combined PET and fMRI study. Neuropsychological Rehabilitation 15:290–307

Owen AM, Coleman MR, Menon DK, Berry EL, Johnsrude IS, Rodd JM, Davis MH, Pickard JD.(2005) Using a hierarchical approach to investigate residual auditory cognition in persistent vegetative state. Prog Brain Res. 150:457–71.

Plum, F., Posner, J. (1982) Diagnosis of Stupor and Coma. F.A. Davis and Company. New York..

Repucci, M.A., Schiff, N.D., and Victor, J.D. (2001) General strategy for hierarchical decomposition of multivariate time series: implications for temporal lobe seizures. Annals of Biomedical Engineering 29(12):1135–49.

Rowe DL, Robinson PA, Rennie CJ. (2004) Estimation of neurophysiological parameters from the waking EEG using a biophysical model of brain dynamics. 231(3):413–33.

Schiff, N.D. and Plum, F. (2000) The role of arousal and 'gating' systems in the neurology of impaired consciousness. Journal of Clinical Neurophysiology 17(5): 438–452.

Schiff, N., Ribary, U., Moreno, D., Beattie, B., Kronberg, E., Blasberg, R., Giacino, J., McCagg, C., Fins, J.J., Llinas, R. and Plum, F. (2002). Residual cerebral activity and behavioral fragments in the persistent vegetative state. Brain 125, 1210–1234.

Schiff, ND, Rodriguez-Moreno, D, Kamal, A, Kim, K.H., Giacino, J, Plum, F and Hirsch, J. (2005) fMRI reveals large-scale network activation in minimally conscious patients. Neurology 64: 514–523

Schiff, ND (2005) Modeling the minimally conscious state: measurements of brain function and therapeutic possibilities. Progress in Brain Research 150: 477–497

Sporns O, Tononi G, Kotter R. (2005) The human connectome: a structural description of the human brain. PLoS Comput Biol. 1(4):e42.

Strauss DJ, Ashwal S, Day SM, Shavelle RM.(2000) Life expectancy of children in vegetative and minimally conscious states. 23(4):312–9.

Szelies, B. et al. (1991) Widespread functional effects of discrete thalamic infarction. *Arch Neurol.* 48: 178–182.

Thatcher RW, North DM, Curtin RT, Walker R A, Biver CJ, Gomez JF, Salazar AM (2001) An EEG severity index of traumatic brain injury. Neuropsychiatry Clin Neurosci 13(1):77–87

Thatcher RW, Biver C, McAlaster R, Salazar A. (1998) Biophysical linkage between MRI and EEG coherence in closed head injury. Neuroimage 8(4):307–26

Tononi G. (2004) An information integration theory of consciousness. BMC Neurosci. 5(1):42

Walker MP, Ayre GA, Perry EK, Wesnes K, McKeith IG, Tovee M, Edwardson JA, Ballard CG. (2000) Quantification and characterization of fluctuating cognition in dementia with Lewy bodies and Alzheimer's disease. Neurology 54:1616–1624.

Voss H.U., Ulug A.M., Watts R., Heier L.A., McCandliss B., Kobylarz E., Ballon D., Schiff N.D. (2005) Regional Increases in Diffusion Anisotropy in a Patient with Severe White Matter Atrophy after Traumatic Brain Injury: A Quantitative Diffusion Tensor Imaging Study, ISMRM 2005 abstract

Index